HUMAN FACTORS IN TECHNOLOGY

HUMAN FACTORS IN TECHNOLOGY

HUMAN FACTORS IN TECHNOLOGY

Editors

EDWARD BENNETT

JAMES DEGAN

JOSEPH SPIEGEL

The MITRE Corporation

Prepared for the Human Factors Society

McGRAW-HILL BOOK COMPANY, INC. *New York San Francisco Toronto London*

HUMAN FACTORS IN TECHNOLOGY

 Library of Congress Catalog Card Number 62-16272

04565

PREFACE

For some time we have thought that those scientific and engineering efforts investigating the relation of man to his technological environment should be reviewed more adequately in the professional literature. Even the new interdisciplinary journals such as *Ergonomics*, the *IRE Transactions on Human Factors in Electronics*, and the *Journal of the Human Factors Society*, depending heavily upon spontaneously contributed articles, have not as yet been able to present as wide a technical overview of the field as is necessary.

With this need in mind, in the spring of 1960, when the Human Factors Society invited us to prepare their fourth annual meeting, we were delighted to accept an opportunity to test some of our ideas about presenting a range of topics focused upon the many-faceted relation between man and his technology. We asked an interdisciplinary team of scientists and engineers to work with us in selecting the subject areas that together would provide a thorough overview of the field. We also asked this team to help us in selecting and obtaining the cooperation of an outstanding man in each of these areas who would, as Session Chairman, be responsible for inviting speakers to represent his particular discipline.

The Fourth Annual Meeting of the Human Factors Society was held on September 12–14, 1960, at the Massachusetts Institute of Technology. By all appearances it was a success, for a truly interdisciplinary audience was given an opportunity to learn of the problems and recent achievements in a wide range of human factors areas.

After the meeting we thought that the publication of a more extensive treatment of many of the same subjects might result in a desirable contribution to the permanent interdisciplinary literature. We asked many of the scientists and engineers who had contributed to the meeting to prepare new, expanded, and documented technical papers. We invited new contributors to report on a number of additional subjects. Those who helped plan and chair the meeting agreed to stay on as technical advisers. The goal of this effort was to bring together material of the highest obtainable quality as a representative sample of the wide range

of scientific and engineering activities that study and help mold the interplay of man and his modern technology. The results, in varying degrees of detail, appear following this Preface.

Blame for inadequacies in the final product must rest with us, since the contributors were all willing to rework their material at our request. Credit for accomplishment, on the other hand, belongs to the authors and to the technical advisers, who generously contributed their time and professional knowledge. Our thanks to the Executive Council of the Human Factors Society for its confidence in us, and particularly to Arnold M. Small and Stanley N. Roscoe, past presidents of the society, for their counsel and encouragement. We are also indebted to Sylvia H. Pilsucki and Barbara R. Robbins, who greatly assisted in editing the manuscripts and who handled the many administrative details of this effort.

Edward Bennett
James Degan
Joseph Spiegel

CONTRIBUTORS

J. S. Abma, Battelle Memorial Institute, Columbus, Ohio (CHAPTER 19)

O. S. Adams, Lockheed-Georgia Company, Marietta, Georgia (CHAPTER 3)

D. E. Bass, U. S. Army Research Institute of Environmental Medicine, Natick, Massachusetts (CHAPTER 2)

T. A. Benham, Haverford College, Haverford, Pennsylvania (CHAPTER 22)

E. Bennett, The MITRE Corporation, Bedford, Massachusetts (CHAPTERS 1 and 33)

J. C. Bliss, Stanford Research Institute, Menlo Park, California (CHAPTER 24)

C. L. Bommarito, The Boeing Company, Seattle, Washington (CHAPTER 11)

B. J. Campbell, Automotive Crash Injury Research of Cornell Aeronautical Laboratory, Buffalo, New York (CHAPTER 13)

W. D. Chiles, Aerospace Medical Laboratory, Dayton, Ohio (CHAPTER 3)

C. Clark, The Martin Company, Baltimore, Maryland (CHAPTER 4)

R. R. Coermann, Aerospace Medical Laboratory, Dayton, Ohio (CHAPTER 5)

R. Contini, New York University, New York, New York (Technical Advisor, PART F)

F. S. Cooper, Haskins Laboratories, New York, New York (CHAPTER 20)

J. Degan, The MITRE Corporation, Bedford, Massachusetts (CHAPTER 1)

C. A. Dempsey, C. A. Dempsey Associates, Dayton, Ohio (CHAPTER 10)

D. B. Devoe, Sperry Rand Research Center, Sudbury, Massachusetts (CHAPTER 38; APPENDIX I; APPENDIX II)

R. G. Domey, Harvard School of Public Health, Boston, Massachusetts (CHAPTER 14)

J. K. Dupress, American Foundation for the Blind, New York, New York (CHAPTER 17)

J. B. Elkind, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts (Technical Advisor)

R. Ford, Harvard Medical School, Boston, Massachusetts (CHAPTER 16)

H. Freiberger, Veterans Administration, New York, New York (CHAPTER 18)

G. A. Fry, School of Optometry, Ohio State University, Columbus, Ohio (CHAPTER 36)

R. J. Gibson, The Franklin Institute Laboratories for Research and Development, Philadelphia, Pennsylvania (CHAPTER 23)

H. von Gierke, Aerospace Medical Laboratory, Dayton, Ohio (Technical Advisor, PART B)

L. O. Gilstrap, Adaptronics, Inc., Annandale, Virginia (CHAPTERS 29 and 30)

R. C. Goertz, Argonne National Laboratory, Argonne, Illinois (CHAPTER 27)

W. W. Haythorn, The RAND Corporation, Santa Monica, California (CHAPTER 35)

G. L. Hekhuis, Aerospace Medical Division, Air Force Systems Command, Brooks Air Force Base, Texas (CHAPTER 6)

A. E. Hickey, Entelek, Inc., Newburyport, Massachusetts (Technical Advisor)

C. O. Hopkins, Hughes Aircraft Company, Culver City, California (CHAPTER 34)

J. L. Kobrick, U. S. Army Quartermaster Research and Engineering Command, Natick, Massachusetts (CHAPTER 9)

M. J. Kopac, New York University, New York, New York (CHAPTER 28)

R. J. Lee, Adaptronics, Inc., Annandale, Virginia (CHAPTERS 29 and 30)

J. C. R. Licklider, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts (CHAPTER 39)

E. E. Loebner, hp associates, Palo Alto, California (CHAPTER 32)

R. A. McFarland, Harvard School of Public Health, Boston, Massachusetts (Technical Advisor, PART D; CHAPTERS 12 and 14)

E. B. Magid, Chicago Medical School and Cook County Hospital, Chicago, Illinois (CHAPTER 5)

M. Metfessel, University of Southern California, Los Angeles, California (CHAPTER 21)

P. Metzelaar, Space Technology Laboratories, Los Angeles, California (CHAPTER 31)

E. F. Murphy, Veterans Administration, New York, New York (CHAPTER 18)

G. Nadler, Washington University, St. Louis, Missouri (CHAPTER 37)

R. W. Newman, U. S. Army Quartermaster Research and Engineering Command, Natick, Massachusetts (Technical Advisor, PART C; CHAPTER 8)

M. J. Pedelty, Adaptronics, Inc., Annandale, Virginia (CHAPTER 30)

P. G. Ronco, Tufts University, Medford, Massachusetts (CHAPTER 38; APPENDIX I; APPENDIX II)

J. J. Ryan, University of Minnesota, Minneapolis, Minnesota (CHAPTER 15)

E. V. Saul, Tufts University, Medford, Massachusetts (Technical Advisor; CHAPTER 38; APPENDIX I; APPENDIX II)

J. W. Senders, Bolt Beranek and Newman, Inc., Cambridge, Massachusetts (Technical Advisor)

T. B. Sheridan, Massachusetts Institute of Technology, Cambridge, Massachusetts (Technical Advisor, PART E; CHAPTER 17)

J. Spiegel, The MITRE Corporation, Bedford, Massachusetts (CHAPTER 1)

J. E. Steele, Aerospace Medical Laboratory, Dayton, Ohio (Technical Advisor, PART G)

G. Stone, New York University, New York, New York (CHAPTER 26)

F. A. Webster, 62 Coolidge Avenue, Cambridge, Massachusetts (CHAPTER 25)

D. W. Williams, The MITRE Corporation, Bedford, Massachusetts (CHAPTER 7)

CONTENTS

Part A

INTRODUCTORY COMMENTS

Chapter 1

HUMAN FACTORS IN A TECHNOLOGICAL SOCIETY

Edward Bennett, James Degan,* and Joseph Spiegel**

SCIENTIFIC OR ENGINEERING STUDIES OF HUMAN FACTORS IN TECHNOLOGY are customarily interdisciplinary. They involve identifiable groups of mechanical engineers, electrical engineers, psychologists, anthropologists, industrial designers, and so forth, working separately or together on common problems. As a result, it is almost impossible to impose a narrow context upon this work; it is more than psychology, more than mechanical engineering. If anything, it is science and engineering devoted primarily to the service of man. As such it represents a positive social force of some magnitude, which is perhaps best understood in its social context.

If there were a password with which to enter the realm of human factors in technology, perhaps the word would be "compatibility." Scientific research and engineering studies of human factors in technology seek to realize greater recognition and understanding of man's characteristics, needs, abilities, and limitations when the procedures and products of technology are being designed. The various parts of this book present some new approaches in a wide variety of settings to this search for "compatibility" between man and his technical environment.

From the clothing he wears to the automobile he drives, from the processed foods on his table to the filtered air of his weather-conditioned office, a contemporary man of the cities lives in an intricate world of things made or modified by man. Even a modern farmer, as he cultivates the soil with his tractor, his chemical fertilizers, and his mail-order tools, must involve himself with the many objects made by men. Such man-made items, in almost infinite array, are the products of that ever-expanding enterprise called technology. This is the enterprise through which men construct, either by transforming the raw materials and energy of nature into simple, useful forms or by combining these manufactured products into more complex and thus more powerful entities. It is within this technological framework that human factors scientists and engineers play their roles. In a technical society there are innumerable situations

* The MITRE Corporation, Bedford, Mass.

in which the characteristics of man may conflict with the characteristics of his technical products and procedures. From these situations grow many of the problems discussed in the following chapters.

The quest for compatibility between man and technology is seldom simple. There are problems in discovering the characteristics and tolerances of people, especially since human behavior is seldom, if ever, hierarchic. There are also problems associated with measuring and evaluating the stresses that technology imposes upon human behavior. Once the contrast between what is desirable and what is obtainable has been established, inventive engineering must meet the difficult task of bringing the real closer to the ideal. The ultimate goal is pragmatic: greater service to man by mechanism, not the reverse.

Man and his technology are intertwined in many ways. This mutual dependence and the problems it generates cannot be exaggerated, either in extent or in depth. The part Human Factors in Atypical Environments scans the matter of human tolerances under the harsh physical conditions of tropical and arctic living, space flight, and thermonuclear war. The part Human Factors in Workspace and Clothing Design points up the extent to which some of the problems that arise from such environmental stress can be overcome by scientific and engineering efforts. Taken together, these parts emphasize the intensity of the interaction between man as a biological and mechanical organism and the heightened stresses imposed upon him by his advanced technology.

Technology is an important source of the goods that consumers desire. It gives producers employment and compensation, which in turn permits them to consume the products generated by others. It urges inventors, engineers, and scientists to study and modify the world. For everyone technology has some relevance, and yet such relevance does not necessarily imply satisfaction, for seldom do the objects created by man fully gratify man's needs. For example, an automobile is a convenient means of private transportation, but it is also a prime killer of both passengers and pedestrians. In the part Human Factors in Highway Safety some of the research in and the recommendations for the reduction of the dangers of automotive transportation are discussed, with emphasis on the importance of increasing the compatibility between the nature of man and the things that man builds for his own use.

Man can sense his world and react to it only imperfectly. As a result, he builds aids to increase his ability to sense and to act. His desire for greater adequacy in these areas accounts for the host of products that extend the power of his eyes, ears, hands, and feet. If the normal faculties of his eyes or arms are injured or lost, his need for aid is compounded. In the parts Human Factors in Sensory Supplementation and Human Factors in Manipulation and Manual Extension, studies of ways by which

to extend man's limited faculties are examined in some detail. These studies illustrate clearly how intimately man and technology are related.

When confronted by the question of whether man and mechanism are more similar than different, one may seek evidence in machines that can act or think as people do. Perhaps through the investigation of activities that are common to human beings and machines the similarities between man and his mechanisms may be better understood. Certainly one can already see that much of man's present burden of work will be relieved as mechanisms are developed that see and hear for man, walk, grasp, and talk for him, and eventually think, even if in a rudimentary fashion. The current search for sophisticated mechanisms to serve man's needs is reflected in the part Human Factors in the Mechanization of Human Functions, which discusses the first tentative successes in the design of ingenious mechanisms capable of helping man with some of his higher perceptual and mental functions.

Where is it all going? What can we expect of this active search for greater congruity between man and his ever-growing technology? With a constantly accelerating scientific and technological base, and with some hope of compatibility between things and people, the growth of a highly rewarding, although dangerous and complex, technological society of computers, automated factories, space flight, thermonuclear weapons, and ballistic missiles may now be seen on the horizon. Within such a society, we may assume that human factors scientists and engineers will play heightened parts. There will be many biological and physical problems to overcome. There will be equal, if not greater, social and emotional ones.

There is a cycle that, once begun, precipitates a continually expanding rush of technological development and introduces ever-increasing complexity and challenge to the interplay of human factors in technology. As long as men are motivated, they will develop new products and the technological means for their production. With time, the invention of new products and new means for production generates an economic surplus, which, in turn, encourages the growth of both national and international trade. Such increased trade encourages men and organizations to improve their techniques and their products in order to raise their economic surplus available for profitable exchange.

This cycle of increasing invention and trade produces still other cyclic effects, each one of which sets a direction for changes that eventually take place. Trade leads to the growth of the various commercial organizations needed to handle the negotiation, transportation, and payment involved. As these commercial organizations multiply, they tend to cluster together in order to facilitate their joint ventures and communication. Thus, just as trade increases commercial inventiveness, it also

increases centralization and it formalizes schemes for administering the movement of products and funds to and from the market place.

Slowly but steadily the isolated cities of a nation give way to centralized metropolises. People become more dependent upon each other as each becomes specialized in the manufacture of the few products that he can make well and economically. Secondary organizations for transportation, communication, and finance grow. The need for delivery men, messengers, and individual moneylenders is replaced by the need for trucking companies, telephone and telegraph systems, and international banks. These metropolises continually expand, for in these giant cities conditions are ideal for the efficient exchange of the products of technology. Communication and transportation distances are short. Intermingling of diverse specialists is easy. The centers of production, the factories, and the centers of consumption, the stores, are close to each other and to the men who must both produce and consume.

Conditions in the cities also speed the pace of technical growth and innovation. Mass advertising and the social pressures that come with close living encourage consumption. New and effective production techniques are rapidly communicated and installed. Potential sources of risk capital and potential creators of industrial organizations are accessible to each other. Thus a secondary cycle forms as increasing technology pushes urbanism and urbanism increases the rate of technological growth.

As a result, human factors problems associated with congested urbanism rise. The need for properly designed urban and suburban transportation facilities becomes more vital as congestion blocks the flow of automobile travel through urban streets. Improved designs for buses and intraurban trains, new strategies for the pleasant movement by land or air of people from home to work and back again, superior procedures for the efficient use of time spent in transit, and better control of the noxious fumes and the waste and litter by-products of such transportation are all required. With the increasing emphasis upon high speed and automatically controlled vehicles, problems of both safety and comfort should continue to be challenging.

As the value of metropolitan land increases, crowded living becomes economically necessary. More and larger skyscrapers increase the congestion of local sidewalks, shopping centers, and entertainment facilities. New room arrangements, furniture styles, building designs, and neighborhood configurations are called for to increase the human satisfaction associated with living in relatively compact homes or apartments. Needed also are radical alterations in the structure of offices, supermarkets, theaters, and other places of mass traffic and congregation.

In a technological society, other human factors problems are created as many new, often alien, inventions appear in rapid succession. The

human factors problems that develop with the arrival of a new technology are generally similar in their emotional and social character whether they arise from the impingement of modern Western technology upon a simple culture, the impact of scientific agricultural methods on a rural community, or the anxieties of a sophisticated city confronted for the first time by atomic energy and its implications of cheap power on the one hand and thermonuclear warfare on the other. The painful impact of a new technology is fairly common. The railroad that pushed the industrial life of Quito, the capital of Ecuador, into the small villages of the Andes brought beef, cheese, bread, and milk, but it also brought alcoholism, moral disintegration, and a rise in land prices. The sudden introduction of automated machinery into contemporary American factories generates discontent, strikes, and slowdowns even today.

Thermonuclear power, space travel, and ballistic-missile warfare all have begun, and will continue, to exert their impact on contemporary technological society. Large cities now can be destroyed completely without sufficient warning for probable survival. Vast distances can be traveled in short times. The weapons of aggression far outstrip the weapons of defense. With all this, traditional concepts of distance, time, and danger are uprooted, and at this point human factors problems of a social and emotional nature arise. Designs for thermonuclear-attack shelters not only must guarantee physical survival but must help supply the emotional and social support that will be essential in maintaining people during the reconstitution to follow. Supersonic aircraft and commercial space vehicles will have to overcome passenger anxieties that are evident even now with less rapid air transportation. The design of atomic reactors for the peaceful generation of electrical power must include characteristics that will alleviate the fears of people living close by.

There is still another class of human factors problems that at present, in elementary form, appears in contemporary advertising and merchandising. These problems revolve around the need to encourage the smooth and comfortable adoption of new techniques and products, especially when such new things arrive in quick succession. At least two sets of forces are involved in the adoption of a new product or procedure. The product or procedure must first be gradually diffused from its point of origin outward and then gradually accepted. The diffusion forces reflect the human factors efforts of the producers, who must convey notions concerning the nature and value of their contribution; adoption forces reflect the cultural, social, and personal situation of the potential consumer. Once a product has been transported, described, and advertised to a community, the individual members of the community then pass through stages in accepting it. At least five such stages in this process of adoption, each with its own human factors implications, can be identified:

Awareness. When a person has discovered the existence of a product or practice but is not necessarily motivated to try it or to seek additional information about it

Information. When curiosity, interest, and personal concern have encouraged the person to seek general information about the product or practice in order to relate it to his own circumstances

Application. When the person, having weighed the advantages and disadvantages, considers using the product or practice but has not actually tried it

Trial. When a limited-scale application of the innovation has been attempted

Adoption. When the trial proves satisfactory enough to encourage continued use

Since there are many individual differences influencing the rates of adoption, the time distribution of this adoption process usually forms a bell-shaped curve. Or if the awareness, trial, and adoption parts are separated out, these three stages form a series of bell-shaped curves. The

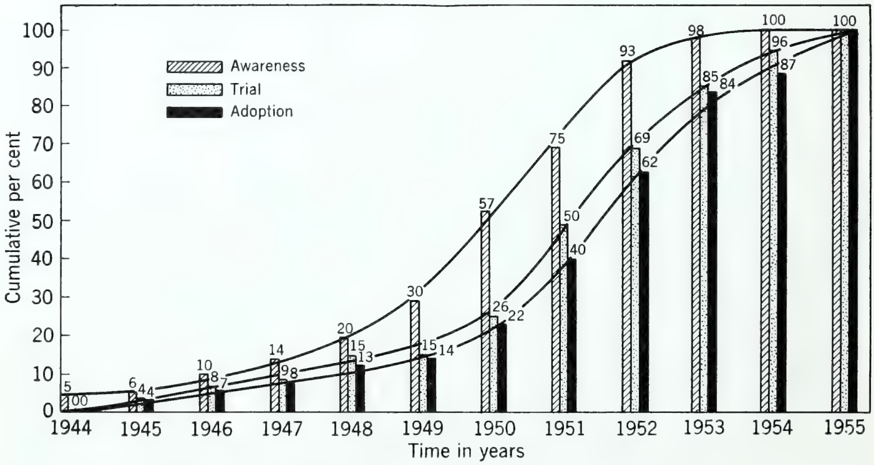


FIG. 1-1 *The adoption of antibiotic pig feed by year in an Iowa community. One hundred forty-eight central Iowa farmers were questioned about their use of antibiotic feed that reduced sickness and death losses while increasing rate of weight gain. The feed was available from most commercial feed companies by 1949. (After G. M. Beal and E. M. Rogers, The Adoption of Two Farm Practices in a Central Iowa Community, State Univ. Rept., Ames, Iowa, 1960.)*

cumulative-effect curves, as in the example of Figure 1-1, reflect the gradual nature of the process.

These adoption curves change with varying conditions. The shape of the *awareness* curve reflects the extent to which the producers have fed

extensive information into the popular channels of communication. These channels include both the formal and the informal channels of personal communication, such as local meetings and small social gatherings, and the popular mass-communication channels, such as magazines, radio, and television. The awareness curve also reflects the degree to which members of the community actually use these available sources of information, that is, how frequently they are motivated to go to meetings, have social affairs, and listen to the radio or watch television. The *trial* curve falls close to the awareness curve either when the diffused information is made persuasive or when the community is technologically liberal and therefore susceptible. In one case the advertising campaign may be particularly insightful, whereas, in another case, people merely may be inclined to try whatever is suggested to them. The *adoption* curve lies close to the trial curve when the new product or technique has a clear preponderance of human factors assets over liabilities, the people have the necessary ability to learn how to use it, the technologically conservative segment of the community is weak, or the actual need for the innovation is considerable.

In many respects the process of adopting new products and procedures reflects the success or failure of human factors efforts up and down the line. Adequate anthropological, biological, and sociological design data, properly human-engineered products, and awareness of consumer preferences and motivation all hasten the adoption of a new technology.

When any newly introduced technology is radically different from the old and is also close to a vital aspect of people's lives, it brings with its adoption new social organizations, new hierarchies for gaining power, new sources of reward, and new values. This is particularly the case with the arrival of novel or radical means for war, transportation, and communication. The machine gun, dynamite, the ballistic missile, the thermonuclear bomb, the railroad train, the automobile, the airplane, radio, television, and the telephone have each exerted its own intense influence.

Nevertheless, even such powerful technical influences are gradually adopted and take their places along with the other aspects of a highly technical environment. The people, the organizations, and the products all balance eventually, and life once again becomes predictable. In fact, once a society has become heavily technical, rapid alterations in technology seldom generate sudden changes in community life. Rather, evolutionary social modifications take place over the years as new and improved technical procedures are introduced.

These gradual changes mirror basic changes in the nature of the relation between man and his technology. An ever-greater proportion of workers have an opportunity to perform the skilled analysis involved in

designing and developing the machines that take over the routine work of older generations. Each phase in the growth of technology results in an increasing need for engineers and scientists to help invent and design laborsaving machines and manufacturing procedures, for skilled toolmakers to build new equipments, and for skilled technicians to install and repair these equipments. There is a decreasing need for semiskilled machine operators and unskilled laborers. The usual effect is that of forcing the range of technical opportunities higher up the intellectual scale. Young, bright workers have improved opportunities. The older, the less intelligent, and the less adaptable workers suffer.

In addition, as factories increase in size, complexity, and automation, an increasing need arises for administrative help to maintain the orderly flow of materials and orders to the plant, to keep track of the processes within the plant, and to arrange for the movement of products out of it. There is an increase in clerical jobs for handling the enlarged burden of paper work. Managers to plan and supervise the organization multiply. The expanded transportation organizations create still more administrative, technical, and clerical jobs. The increased need for communication also creates new technical-job opportunities. For any one technical skill that is made obsolete a dozen new and more complex ones are required.

With this technical growth, skilled labor is often in short supply, and the financial return for such labor is high. Jobs for the skilled are not too difficult to find, and income is adequate or better. People tend to relax their efforts as their situation improves. For the people involved in production, pressures tend to level off.

Therefore, under these conditions the major human factors frontiers become those of automation, for the environment encourages technological invention, and its greatest need in this respect is for machines that can do more and more of the work that man does not want to do. The human factors problems are primarily those of developing machines to take over the more complicated parts of man's job and developing complex systems of men and mechanisms to work effectively together. Therefore, better understanding of how man does what he does, how he moves, holds, talks, judges, and decides, is needed. Also needed is a clearer view of how man and machine may best communicate, what language, what symbols, and what speed are involved.

Mechanization follows the normal pattern of technological growth, each step of which transfers some portion of the technical process from man to a newly developed machine. The rate of transfer hinges upon the ability of men to feed the technology with new inventions. Past progress will have mechanized the simpler jobs of human beings. The mechanization of remaining areas of human activity, however, requires a level of inventiveness that may eventually equal or surpass the skill of even a

highly creative and sophisticated society. As a result, the last stages of mechanization, which we think of as automation involving adaptive computer control, are still slow and hesitant.

The jobs that are left for men to do in a highly automated society are the jobs which they enjoy or those which machines cannot do economically or at all. This kind of work is mainly process planning and system engineering. The jobs tend to include unique and unpredictable problems requiring some of man's natural flexibility. They are mainly jobs that require an understanding of the changing pattern of numerous factors. In most cases these tasks are not hierarchic and, therefore, cannot be reduced to simple and relatively independent subordinate units. They are jobs that commonly involve creativity and ingenuity.

Because of the change that takes place in job emphasis, industrial human factors efforts that at present tend to focus upon bettering man's ability to do simple, routine factory or clerical tasks may be gradually redirected. The new need is for improved procedures to handle the creative problem-solving jobs, especially those associated with the subtleties of consumer behavior, interpersonal relations, and complex system design and engineering. Some of the tools that human factors specialists are now using in this kind of effort are discussed in the part Human Factors Methods and Procedures. However, the complex systems of the future may well have to be built by the kind of "system system" that is outlined in the concluding part of this book.

Part B

HUMAN FACTORS IN ATYPICAL ENVIRONMENTS

Chapter 2

TEMPERATURE REGULATION IN MAN

*David E. Bass**

THE ABILITY TO MAINTAIN A WARM-BLOOD STATE—HOMEOTHERMY—CONFERS upon man obvious advantages in terms of freedom to migrate over a wide range of environments while maintaining normal alertness and effectiveness. This homeothermic state, however, imposes an appreciable cost on the physiological economy, for it requires its own regulatory apparatus and constant maintenance. Even a temporary lapse of maintenance imperils not only normal effectiveness but life itself.

It should be realized at the outset that the term "normal" body temperature may be confusing. The familiar oral temperature of 98.6°F is not the universal constant that it is sometimes considered to be by hypochondriacs and anxious parents; it represents only one of many body temperatures. Nevertheless, it is a clinically useful approximation of the temperature of the body "core," which contains the vital organs. A better assessment is rectal temperature, which is 1°F higher than oral temperature. When physiologists refer to "core" temperature, they usually mean rectal temperature [7]. Not all the core's components are at the same temperature; for example, the liver is usually 1 to 2°F warmer than the kidneys. Furthermore, a healthy person can function with elevated core temperatures resulting from exercise and other situations; for instance, athletes running distance events have displayed rectal temperatures as high as 105°F with no discomfort. Finally, there is a well-defined diurnal rhythm of core temperatures, with a range of up to 3°F, the lowest value occurring in the early morning and the highest in the evening. This diurnal cycle is largely the result of daily activity and ingestion of food.

Outward from the core to the body's surface there normally exists a temperature gradient, the skin surfaces being coolest, as shown in Figure 2-1. The noncore part of the body is usually termed the "shell"; it comprises mainly the skin and parts of the subcutaneous layers. In thermo-

* Cellular Medicine Division, U.S. Army Research Institute of Environmental Medicine, Natick, Mass.

This chapter represents excerpts from a report submitted to the Secretary of the Army by the author on his activities during tenure of a Secretary of the Army Research and Study Award.

physiology the concept of core and shell has been useful for understanding body-heat exchanges. The core is considered to be that part of the body in which heat is produced and temperature is closely regulated. The shell is that part in which heat exchanges are modified in support of core regulation. For example, portions of the skin (e.g., the hands and feet) can cool to freezing while the core temperature is 99°F. Even in comfortable conditions the mean skin temperature is usually 8 to 10°F cooler than the core. An important aspect of temperature regulation is the alteration of temperature gradients by physiological and cultural adjustments, between core and skin and between skin and the surroundings, which modifies the amount of heat lost according to the needs of the body.

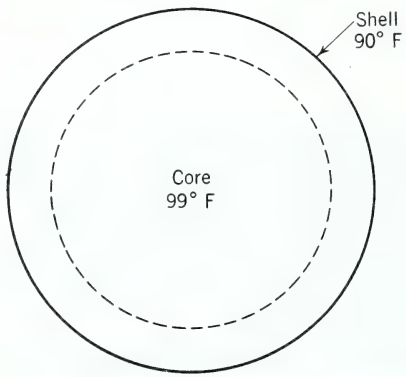


FIG. 2-1 *Diagrammatic representation of body core and shell. In the cold the shell is greatly thickened and may attain surface temperatures down to freezing. In the heat the shell becomes very thin because of vasodilatation and may attain temperatures higher than the core temperature.*

Although there is no single, fixed body temperature, there is a relatively narrow range of core temperatures that must be maintained for normal physiological function. This range is 97 to 102°F for most people. Although athletes and acclimatized workers in hot industrial environments may function well with temperatures as high as 104 to 105°F, such cases are exceptional. It is interesting to compare the range of temperature for optimal function with that compatible with life. Although man can survive core cooling to 77°F, his margin of safety on the hot side is much smaller. When his core temperature reaches 106°F—an increase of only about 7°F—he is in danger of heat stroke, with the possibility of brain damage and even death. Death usually results when core temperature reaches 110 to 112°F.

FACTORS IN HEAT BALANCE

Man's thermal equilibrium is simply the result of a balance between heat gain and heat loss (Figure 2-2). It may be expressed by a simple equation:

$$M - E \pm C \pm R = 0$$

where M = heat of metabolism (always positive)

E = heat lost via evaporation (always negative)

$$\left.\begin{array}{l} C = \text{convection} \\ R = \text{radiation} \end{array}\right\} \text{may be either positive or negative}$$

When equilibrium is temporarily upset, as in exercise or in cold exposure, the right side of the equation differs from zero. Conduction is not included in the above equation, since it is not subject to regulation and depends solely on physical contact, e.g., with a chair or the ground.

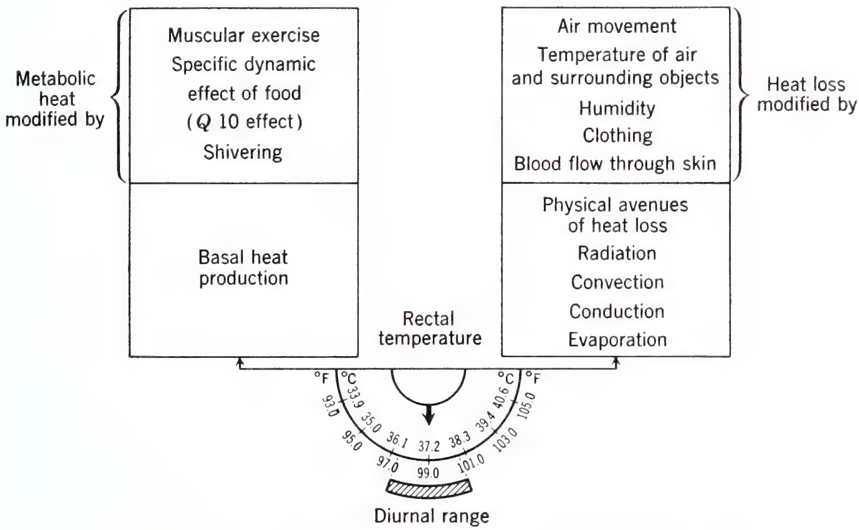


FIG. 2-2 *Factors that influence heat balance in man. (After E. R. Buskirk and D. E. Bass, Climate and Exercise, in "Science and Medicine of Exercise and Sports," ed. Warren R. Johnson, pp. 311–338, Harper and Brothers, New York, 1960. By permission of the publishers.)*

Some knowledge of the factors in the heat balance is required for a clear understanding of physiological thermoregulation. These will be discussed briefly at this time.

Metabolic Heat

Normally, the major source of heat gain to the body is metabolism. Metabolism, in its simplest meaning, is synonymous with "change." In the context of this discussion, metabolism refers to the sum total of the chemical and physical reactions in the body. It is not always correct to equate "metabolism" with "heat production," since some of the energy of metabolism is converted into work or supports chemical reactions with varying degrees of efficiency. The more exact expressions are "metabolic

heat" or "heat of metabolism." Heat, nevertheless, is the largest by-product of metabolism. The efficiency of metabolism is proportional to the percentage of nonheat energy made available to the body; in other words, the greater the proportion of heat produced, the less efficient is the metabolism. The heat, however, is useful in keeping the body warm. All metabolic reactions occur in the cells, and the immediate source of energy is "high-energy" compounds previously synthesized and stored therein. These stores must be periodically replenished; this is accomplished by food eaten.

If food is burned (i.e., completely oxidized), as in a bomb calorimeter, to form carbon dioxide and water, it is possible to measure accurately the energy available in the food as heat. By this method the following values have been found for the three main classes of food: fat 9.3 kcal/g; protein 5.6 kcal/g; carbohydrate 4.2 kcal/g.[1] The body, too, oxidizes its energy sources, but at a slower rate and at a lower temperature than that obtained when food is burned in a flame.

The efficiency of the body in terms of energy utilization depends on the activity and type of work done by the various tissues. Although some isolated cellular enzyme systems have been shown to have efficiencies of up to 50 per cent in the test tube [2], the overall efficiency of man is much lower, ranging from almost zero in the resting state to a maximum of approximately 25 per cent when external work is done; the latter figure compares favorably with many engines. When a man is reclining quietly in the morning before breakfast, his heat production is approximately 70 kcal/hr. This is termed the basal metabolic rate; it is the lowest heat production exhibited by the healthy, awake man. The basal heat production is by no means inconsiderable; it is, for example, enough to heat a quart of water from room temperature to the boiling point; it is approximately equal to the heat produced by a 100-watt electric-light bulb. It is large enough to impose a requirement for heat dissipation at comfortable room temperatures. This is illustrated by the fact that, when the body's ability to dissipate heat is impaired, death from hyperthermia can occur at a room temperature of 75°F.[2]

It is of interest to compare the contribution of the various parts of the body to the total heat production at rest and during severe muscular work. When the body is at rest, approximately 58 per cent of the heat comes from tissues that total only 8 per cent of the body weight, the liver, kidneys, brain, and heart. The skeletal muscles, which account for 50

[1] A kilogram-calorie (kcal) is the amount of heat required to raise 1 kilogram of water 1°C from 15 to 16°C at sea level.

[2] Hyperthermia is a higher than normal deep-body temperature. A lower than normal deep-body temperature is called "hypothermia."

per cent of the body weight, contribute only 20 per cent of the basal heat production. However, during severe exercise the heat production can increase to 1400 kcal/hr, or 20 times the basal rate, and in this situation the skeletal muscles account for most of the increased heat production.

Factors Affecting Metabolism

Several factors affect the body's heat production. These are (1) body size, (2) basal metabolic rate, (3) growth, (4) food, (5) muscular exercise, (6) environmental temperature, and (7) hyperthermia.

Body Size. It is axiomatic that larger individuals have greater energy exchanges than smaller persons. The range of basal heat production in a group of young men may vary as much as 40 per cent of the mean, with, for example, a small man producing 56 kcal/hr and a large man 88 kcal/hr; both men would be in thermal equilibrium and would have similar body temperatures. The increase in basal heat production with body size is not proportional to any single measure of size, although basal heat production does vary with weight to the two-thirds or three-fourths power.

For many years it was thought that, since all homeotherms have approximately the same core temperature and since heat loss occurs mainly at the body surface, homeotherms' metabolic heat must be produced in proportion to heat loss, i.e., to the surface area of the body. DuBois derived the following equation for calculating surface area from height and weight [8]:

$$\text{Surface area (m}^2\text{)} = \text{weight (kg)}^{0.425} \times \text{height (cm)}^{0.725} \times 71.84$$

Most of the data in the literature express heat production in kilogram-calories per square meter of surface area. This has proved to be empirically useful in comparing results obtained with individuals of different body sizes. Thus, in the example cited above, the small and large man might have identical rates of heat production per square meter of surface area. From a physiological viewpoint, however, the rationale for using surface area as a reference standard is rather weak. For example, the body does not lose heat from its entire surface, but from probably 80 per cent or less, depending on posture; also, extremes of fatness or leanness are not adequately reflected in surface area.

In recent years the concept of the mass of metabolically active tissue as a determinant of basal heat production has gained wide acceptance [3, 4]. This concept states that heat is produced only in the metabolically active, fat-free tissues of the body and that therefore the basal heat production is proportional to the mass of this tissue. Thus, two men with

identical fat-free masses may produce the same basal heat even though they differ in weight by as much as 20 lb because of different amounts of body fat.

Basal Metabolic Rate (BMR). When body size is taken into account, the variability in BMR is considerably reduced but not eliminated. The residual variability is due to individual differences in thyroid and other glandular activity. In a large population approximately 80 per cent would display BMRs ranging ± 10 per cent of the group mean, and 90 per cent would fall within ± 15 per cent. BMR is increased with increased thyroid activity and is decreased with starvation and decreased thyroid function.

Growth. The laying down of new tissue requires extra energy, and a growing youth has a higher BMR for his size than an adult.

Food. The ingestion of food is accompanied by an increase in heat production. This is in excess of the energy value of the food and is termed the specific dynamic action (SDA) of the food. Protein has the highest SDA, providing 25 to 30 per cent excess heat; fat and carbohydrates have smaller SDAs. This excess heat is not accounted for by increased digestive glandular and intestinal activity; rather, it is probably the result of intermediary metabolism of the food.

Environmental Temperature. Ambient heat does not greatly affect basal or resting metabolism. Cold, on the other hand, has a marked effect by way of causing shivering. At 60°F a nude man increases his resting heat production by 20 to 25 per cent. In colder situations shivering can increase "resting" heat production by as much as four times that observed in comfortable environments.

Hyperthermia. It is well known that the rate of chemical reactions is approximately doubled for each 10°C rise in temperature. It has been shown that metabolism increases approximately 7 per cent for each Fahrenheit-degree rise in core temperature [13]. The undesirable aspects of this effect during fever are obvious.

Physical Factors

The physical avenues of heat exchange are radiation, convection, evaporation, and conduction. Although these are defined in any physics textbook, the first three merit brief discussion at this point.

Radiation. This form of heat exchange occurs at the surface of the body and is a function only of the temperatures and nature of the radiating surfaces. The radiation from the body to cooler surfaces occurs in the infrared region of the spectrum, and within this range the skin acts almost as a perfect black body; i.e., it is an excellent radiator. The skin is also an excellent absorber of infrared radiation. Skin color does not affect these qualities, since the radiation is in the invisible part of the

spectrum. Skin color does, however, affect absorption and reflection of visible radiation from the sun; thus white skin reflects 30 to 40 per cent of solar radiation, whereas Negro skin reflects less than 18 per cent.

Convection. Little can be added to the usual definition. It should be pointed out, however, that convection is an important avenue of redistribution of heat within the body, via the circulating blood. This serves two useful purposes: (1) it brings metabolic heat to the surface of the body, whence it is dissipated; (2) it rapidly transports heat from active tissues, e.g., the muscles, heart, and liver, and thus prevents the development of "hot spots" during periods of great activity.

Evaporation. It should be pointed out that, even in the absence of visible sweating, there is a continuous loss of water vapor from the body. There are two sources of water-vapor loss: (1) expired air, which is almost completely saturated with water vapor at normal oral temperature; (2) the intact skin, by way of diffusion. This vapor loss is termed "insensible water loss," since it is not normally visible. In cold weather the familiar foggy breath is evidence of this loss from the respiratory tract, and at minus 40°F the loss from the intact skin may be seen as a fog rising from briefly exposed hands. The total insensible water loss is not negligible, amounting to approximately 1 qt of water daily; approximately 25 per cent of basal heat production is dissipated in the formation of insensible water. The magnitude of this loss is not modified in response to thermoregulatory requirements; this poses a potential disadvantage in cold weather. The major source of evaporative heat loss in exercise and heat stress is, of course, from sweat, which *is* secreted in response to thermoregulatory needs.

The proportion of heat lost by the various physical avenues varies with environment and activity. In the cold (in the absence of wind and clothing), radiation is the major source of heat loss; in the heat, evaporative heat loss predominates. A nude man at rest in a comfortable environment with low air movement loses 60 to 65 per cent of his heat by radiation and 20 to 30 per cent by evaporation. In a cold environment with significant air movement, convective heat losses assume a major role. In the heat, wind may cause a convective heat gain if air temperature is higher than skin temperature; however in dry heat, this effect is greatly overshadowed by the enhanced evaporative cooling that results.

PHYSIOLOGICAL REGULATION OF BODY TEMPERATURES

Almost 100 years ago the great French physiologist, Claude Bernard, set the stage for modern physiology when he pointed out that the blood and lymph that bathe the body cells constitute an internal environment

that must be "fixed" if life is to continue [5]. In the 1930s Cannon extended this principle to include a consideration of the mechanisms that regulate the "fluid matrix" of the body. Cannon's conclusion was that normal body function during "extensive changes in the outer world" depends on maintenance of a steady state in the fluid matrix and that this steady state is the result of many regulatory physiological mechanisms acting in concert. To this important unifying concept Cannon gave the name "homeostasis" [6]. This steady state is brought about by an exquisitely complex system of physiological regulations involving all organs of the body, many with their own homeostatic controls. Thermal regulation is one of the most readily observed facets of homeostasis.

Acute Responses to Heat and Cold

A description of the physiological responses to acute heat and cold stress provides a logical starting point for considering temperature regulation. A nude man reclining quietly in an environment at 85 to 88°F with low relative humidity is in a thermally "neutral" state; i.e., he can remain thus indefinitely without becoming uncomfortably warm or chilled. In a still atmosphere, convective heat loss from the skin surface is negligible; so also is conductive loss, since air is a poor conductor. Most of the heat is lost through radiation to the surroundings and through evaporation of insensible water. Within the body, heat is transported to the surface largely by convection (blood flow) and to a much smaller extent by conduction through the tissues. In this "neutral" state, the strain on thermoregulatory mechanisms is minimal.

If room temperature is increased slightly, certain physiological responses occur. The skin blood vessels dilate, heart rate increases, and there is an increase in the flow of blood to the skin. In terms of the heat-balance equation, these responses accomplish the following: the increased skin blood flow brings a greater amount of metabolic heat from the core to the surface; this effect increases the skin temperature; as a result, radiative heat loss is also increased, and a new steady state is attained. If ambient temperature is further elevated, the body's ability to maintain thermal balance by radiation alone is soon overcome. Indeed, as the temperature of the surroundings exceeded that of the skin, the body would gain heat from the environment if it did not have in reserve another source of heat loss. This reserve is, of course, the elaboration of sweat. The sweat glands provide the major physiological defense against overheating. The body can secrete over 3 liters of sweat in 1 hr of strenuous exercise in hot weather; this represents a very large avenue of heat loss, since the evaporation of each liter of sweat water removes approximately 580 kcal of heat from the body. Thus 3 liters of sweat, if completely evaporated, can rid the body of an amount of heat equivalent to 24 times the basal

heat production. It should be emphasized that only sweat that evaporates possesses "cooling power"; sweat rolling off the body contributes little to heat loss. This explains the oppressiveness of "muggy" days, when evaporation is hampered by high humidity.

If the neutral environment is cooled, the first physiological response is a constriction of the skin blood vessels. This serves to reduce skin blood flow and therefore to keep warm blood away from the surface. The skin is cooler, and radiation from the body is reduced. It is common to observe a slight increase in core temperature, since metabolic heat is now distributed in a smaller core contained in a thicker shell. In addition to generalized vasoconstriction, certain interesting and important changes in local blood flow occur; for example, the blood flow through the fingers and toes is disproportionately reduced because of the activity of vascular shunts in these areas. In cold men, the finger blood flow can be reduced to as little as 2 per cent of precold values, and thus the fingers are transformed into essentially bloodless masses of tissue. Since the extremities normally have a high blood flow relative to their volume, this selectively large reduction in the cold is an important mechanism for reducing heat loss, albeit at the risk of frostbite.

As is the case with heat stress, changes in vasomotor activity confer only limited protection against cold. (The author has observed nude men to complain of chilliness after 2 hr at 83°F.) As a man becomes colder, his most important protective response—shivering—is activated. Shivering is involuntary, uncoordinated muscular activity and adds to metabolic heat production; it may be manifested as slight muscle tension during mild chilling or as violent, uncontrollable spasms when the man is very cold. Although the most violent shivering of which man is capable is not so effective for rewarming as is severe exercise, it can, nevertheless, add up to 350 kcal/hr. It has been reported that shivering can retard or prevent further heat loss in an already cold man but will not replace the heat already lost from the body [13].

In summary, the body's first line of defense against both heat and cold is modification of heat transport to the skin by means of vasomotor activity. This mechanism has limited protective capacity, as illustrated by the insulative values of the skin in different environments; these values range from 0.1 to 0.8 clo units for maximal vasodilation and maximal constriction, respectively.[3] The body's major defenses, sweating and shivering, provide remarkably great buffering action against heat and cold.

[3] A clo is the amount of insulation necessary to maintain comfort and a mean skin temperature of 92°F at an ambient temperature of 70°F with air movement not over 10 ft/min and relative humidity less than 50 per cent. Metabolic heat production in these conditions equals 50 kcal/m^2/hr. The ordinary business suit has an insulation value of approximately 1 clo.

Physiological Control Mechanisms

An important characteristic of most homeostatic mechanisms is their reflex nature; that is, they operate speedily and at the subconscious level. Thermoregulation is this type of mechanism; it operates during natural sleep, and it controls body temperature even in animals whose cerebral cortices have been removed. The physiology of temperature regulation is, in its simplest outline, as follows: There exists in the base of the brain (the hypothalamus) a thermoregulatory center. This center has two distinct areas: the anterior (forward) area contains a group of nerve cells that control the avenues of heat loss (vasodilatation and sweat production); the posterior area, slightly to the rear, is composed of nerve cells that control heat conservation (vasoconstriction and shivering). These two subcenters, acting in coordination, provide "thermostatic" control of the body core. Associated with these centers are highly specialized nerve cells that are sensitive to very small changes in temperature. Also, these centers are connected by nerve pathways of varying complexity to each other and to hot and cold receptors and blood vessels in the skin, sweat glands, skeletal muscles, and higher brain centers. Two types of afferent, i.e., incoming, messages reach the thermoregulatory centers:[4] (1) direct temperature impulses, by way of the brain blood supply, and (2) nerve impulses, which originate in the cutaneous hot and cold receptors. The nerve impulses pass along afferent nerves to the spinal cord and then up the cord to the midbrain, where they connect with two further pathways, one going to the hypothalamus, the other continuing up to the sensory cortex, where conscious thermal sensations are perceived. The hypothalamic center integrates these afferent impulses and the blood temperature and sends efferent impulses along diverse pathways down the spinal cord and out to the sweat glands, cutaneous blood vessels, and skeletal muscles, in which structures are evoked for the appropriate responses already described. Although the hypothalamic control center has been likened to a thermostat, it differs from the familiar household thermostat in that it is not an ON-OFF type of control and it responds to a continuous input of several types of signals.

The foregoing summary indicates that the nervous system plays a central, prepotent role in regulation of body temperature. Those aspects of this system which are most involved are (1) thermal receptors, (2) the autonomic nervous system, and (3) the hypothalamus.

Thermal Receptors. There is no doubt that the skin contains two types of receptors sensitive to heat and cold, respectively. Painstaking

[4] "Afferent" means centripetal, i.e., conveying impulses or stimuli from the periphery to the brain (e.g., nerve impulses from the skin to the central nervous system). "Efferent" refers to the conveying of impulses away from a nerve center, as from the brain to the skin.

exploration of the skin has revealed spots that respond selectively to heat and cold. On the forearm, cold spots greatly outnumber warm spots (7:1). There is surprisingly little knowledge, however, concerning the anatomy of these receptors. The warmth receptors are located deeper in the skin than the cold receptors. Hensel estimated the cold receptors to be at a depth of 180 μ [11]; Hendler and Hardy reported a depth of 200 μ for warmth receptors [10].

Normally the receptors initiate continuous volleys of nerve impulses that reach both the conscious centers and the hypothalamus by the nerve pathways previously described. Sensations of warmth or cold vary in intensity with the rates of impulses discharged by the appropriate receptors. Within limits, these sensations exhibit adaptation, i.e., they disappear despite continued stimulus of warmth or cold; this adaptation is commonly experienced when one first enters a warm bath. When small skin areas are stimulated, sensory adaptation can occur between 68 and 104°F. When the entire body area is stimulated, the "neutral" range of mean skin temperatures is 89 to 95°F. When the integrated skin temperature is above 95°F, a sensation of warmth is always present; the threshold for persistent cold sensation is not well established for total skin cooling. Adaptation is affected by the rate of change of temperature and by the number of receptors thus stimulated.

Cold receptors discharge maximally over the temperature range of 59 to 68°F; the total range over which they discharge impulses is 50 to 105°F. Warmth receptors exhibit maximal rates of discharge at 99.5 to 104°F; their range for detectable discharge is 68 to 116°F. When cold receptors are rapidly cooled, they respond initially with an "overshoot," i.e., a burst of impulses, followed by a subsidence to the rate characteristic of the ultimate temperature attained. Rapid warming of cold receptors is accompanied by an "overshoot inhibition" followed by a resumption of the discharge rate associated with the final temperature. Warmth receptors display similar overshoots, but in opposite directions from those of cold receptors.

In summary, the skin contains both warm and cold receptors. The cold receptors are appreciably the more numerous. Both types of receptors continuously send afferent impulses to the brain. The rate at which these signals are sent (and presumably the intensity of thermal sensation) are governed by three factors, (1) the temperature at the receptors, (2) the rate of change of temperature, and (3) the number of receptors involved.

The Autonomic Nervous System. Many organs concerned with homeostasis are autonomic; i.e., they can maintain a certain level of activity without any nerve supply. These organs are those which contain smooth muscle or secretory cells. Examples of the former are the blood vessels (except capillaries), digestive tract, and bladder; the latter include the

sweat glands and the salivary and other digestive glands. The heart muscle, which resembles both smooth and skeletal muscle, can also contract without a nerve supply. A state of physiological anarchy would exist if all these organs functioned *pari passu* without central coordination. Such coordination is supplied by the aptly named "autonomic nervous system." This system consists of two divisions, the sympathetic and the parasympathetic. These divisions are anatomically and functionally distinct, but they act in close harmony—at times stimulating, at times inhibiting—to regulate autonomic function according to the needs of the body.

The sympathetic division originates in nerve cells located along the thoracic and upper lumbar portions of the spinal cord. These cells contain efferent fibers that leave the cord along the ventral spinal nerves and continue to two chains of ganglia (i.e., groups of nerve-cell bodies) situated on either side outside the cord. The fibers form synapses with ganglionic cells, which in turn send postganglionic fibers to "target" organs. The preganglionic fibers discharge tonically[5] even in the basal state, and one fiber serves several ganglionic cells. By this arrangement, the sympathetic nervous system is capable of a diffuse and even mass discharge. This is facilitated by the adrenal medulla, which is an integral part of this system. It receives only preganglionic fibers that stimulate the release of adrenalin and nor-adrenalin into the blood. Circulating adrenalin by itself can produce the effect of a mass sympathetic discharge, since postganglionic sympathetic fibers produce their characteristic effects by release of either adrenalin or nor-adrenalin at the target organs. Thus these fibers are termed "adrenergic." The only exception is the sympathetic innervation of the sweat glands; this will be discussed shortly.

At this point two questions arise: (1) What are the effects of sympathetic stimulation? (2) What would cause a mass sympathetic discharge? Among the important effects are: increased rate and vigor of contraction of the heart; vasoconstriction in the skin, viscera, and kidneys; dilatation of the muscle blood vessels, coronary arteries, and bronchioles; dilatation of the pupils; mobilization of blood glucose. With mass discharge the foregoing all occur simultaneously, and they appear to prepare the body for integrated, lifesaving action in an acutely threatening situation. The effects on the heart prepare it for increased work; the cutaneous, visceral, and kidney constrictions divert blood from areas of quiescence to areas where it will be needed, i.e., to the skeletal muscles, whose blood vessels dilate in anticipation of the greater blood flow, with its accompanying greater oxygen supply and "flushing away" of metabolic

[5] Exceptions are the preganglionic fibers associated with the sweat glands and adrenal medulla.

waste; the coronary arteries are prepared to supply a greater amount of blood to a harder-working heart; the bronchioles are dilated, with a consequent decrease in resistance to respiration. Cannon recognized this coordinated response and appropriately called it the "fight or flight" reaction. He listed several situations that elicit it: fear, rage, pain, asphyxia, heavy muscular work, and exposure to cold [6].

Several centers in the central nervous system regulate the tonic discharge from the sympathetic preganglionic fibers. Centers exist in the spinal cord for the mediation of spinal reflexes, in the medulla for reflex control of cardiac and respiratory activity, and in the hypothalamus, where sympathetic aspects related to temperature regulation and strong emotional states are regulated. In addition, there are ill-defined connections from the higher, conscious centers that can affect sympathetic activity. The sympathetic nervous system is indeed ubiquitous and well organized for emergencies.

The parasympathetic system arises from cranial and sacral nerves and is as varied as the sympathetic system in its distribution and spectrum of effects. It differs from the sympathetic in the following aspects: (1) the parasympathetic ganglia are located close to or actually within their target organs, and as a result there is a restricted distribution of postganglionic fibers; (2) the chemical mediator of parasympathetic activity is not adrenalin but acetylcholine, which is rapidly destroyed by an enzyme (cholinesterase) at the nerve endings; (3) there is no analog of the adrenal medulla that is capable of secreting acetylcholine into the general circulation. The parasympathetic system is, therefore, more discrete in its effects, and its role appears to be directed toward regulating individual-organ activity and conserving body resources for the "long haul." It responds to individual-organ requirements rather than total-body emergencies. It is not surprising, therefore, that there are no well-defined parasympathetic regulatory centers comparable with those which regulate sympathetic activity.

Several organs receive innervation from both divisions of the autonomic nervous system. In such cases, the effects are usually antagonistic; that is, if one division is excitatory, the other will be inhibitory. This is perhaps best illustrated by the heart, which receives parasympathetic stimuli from the vagus nerve and sympathetic innervation from the cardioaccelerator nerve. The heart rate at any time is the resultant of the tonic discharges over these nerves. Parasympathetic stimulation slows the heart; sympathetic stimulation accelerates it. In contrast, parasympathetic nerves to the abdominal viscera stimulate peristalsis, whereas sympathetic stimulation inhibits it. The blood vessels and sweat glands receive only sympathetic fibers. These are excitatory with respect to the skin vessels (vasoconstriction) and sweat glands (sweat secretion) and inhibitory

for muscle vessels (vasodilatation). In addition, there is evidence that some of the sympathetic nerves to muscle vessels produce active dilatation. The thermal sweat glands are unique in that they receive sympathetic fibers that are cholinergic. Sweat glands in the palms and soles are adrenergic. Thus, atropine, which blocks acetylcholine, also inhibits thermal sweating; conversely, acetylcholine or its more stable analog, Mecholyl, can induce thermal sweating when injected locally.

In summary, the autonomic nervous system subserves homeostasis by regulating autonomic function in two ways. It prepares the body for acute emergencies by its diffuse sympathetic division, and it supports individual organ function via the parasympathetic. These two divisions, although dissimilar in their effects on the same organ, are closely coordinated. They both exert their action by way of chemical mediators—adrenalin and nor-adrenalin for sympathetics and acetylcholine for the parasympathetics. The sympathetic system is the more closely related to thermoregulation. Finally, and very importantly, the autonomic nervous system is an efferent motor system, operating upon organs with smooth muscle or secretory cells. It may be considered as the efferent arc of reflexes whose afferent pathways are diverse with regard to modality and location, e.g., temperature and pain from the skin, stretch and chemoreceptors within the cardiovascular, digestive, and respiratory systems, light from outside the body, emotions from higher centers. These afferent pathways are *not* part of the autonomic nervous system.

The Hypothalamus. The hypothalamus contains integrating centers for several important homeostatic functions. Among these are water balance, appetite, food intake, sympathetic activity, and temperature regulation. The cells that control these functions are not well defined, although general areas of control have been mapped through the study of physiological effects of small lesions and electrical stimulation. The heat-dissipative center is anterior; the heat-conservation center is posterior. There is little doubt that these are the major and probably the only thermoregulatory centers in the body. This is demonstrated by the fact that if the brain of an animal is sectioned below the hypothalamus there is complete loss of thermoregulation. If the section is above the posterior hypothalamus, the animal can protect itself against cold but not against heat; such an animal may even die of heat stroke at room temperature because of accumulation of metabolic heat. Finally, local cooling or warming of the hypothalamus elicits corresponding protective responses against cold or heat, irrespective of ambient temperatures.

The Skin Circulation

Blood reaches the skin by way of subcutaneous arteries that divide to form two vascular complexes, one slightly deeper than the other. In

the skin itself the blood flows through vessels in the usual order, i.e., arteries → arterioles → capillaries → venules. The following considerations are pertinent to the role of these vessels in thermoregulation: (1) The cutaneous blood vessels are located only 2 to 3 mm from the skin surface. (2) The vascularity and blood flow are far greater than would be required by an organ with the low metabolic requirements of the skin; it is logical to assume, therefore, that the blood serves an extrametabolic function, which is undoubtedly thermoregulation. (3) The arterioles are invested with a coat of smooth muscle with circularly arranged fibers. This coat possesses an intrinsic tone and receives sympathetic innervation. The venules are similarly innervated and contractile, but to a much lesser extent than the arterioles. (4) The capillaries are thin conduits that, although intrinsically elastic, are noncontractile and receive no nerve supply. They do not normally regulate blood flow.

Thermoregulatory changes in skin blood flow are effected by constriction or dilatation of the arterioles in response to changes in sympathetic tone. In other words, vasoconstriction results from increased sympathetic stimulation; vasodilatation results from a decreased sympathetic discharge. The sympathetic effects are controlled by the hypothalamic centers. Thus the tonic nature of the sympathetic nervous system and of the hypothalamus permits excitation and apparent inhibition without the presence of specific inhibitory innervation.

Blood flow to the skin (and other areas) can be affected by nonnervous factors. For example, the local accumulation of metabolites causes dilatation. Indeed, this is an important mechanism for regulating blood flow to specific organs according to metabolic need. As an organ's metabolism increases, its requirements for oxygen and for dissipation of metabolic waste products increase; the local vasodilatation meets these requirements by increasing blood flow. Injury also causes vasodilatation, presumably by release of histamine, which is a potent dilator.

THE SWEATING MECHANISM

Structure and Distribution of Sweat Glands

Human skin contains two types of sweat organs—the apocrine and the eccrine sweat glands. The apocrine glands are localized mainly in hairy regions, e.g., the armpits and genital areas. Their secretion is thick, containing much cellular matter, and they play no role in thermoregulation; therefore, they will not be considered further. Thermal sweat is produced by the eccrine glands. These are small, tubular coils situated 1.3 to 3 mm below the skin surface; each coil secretes sweat, which pro-

ceeds along a duct that spirals through the upper dermal layers in corkscrew fashion and terminates at the skin surface as the sweat pore. The average volume is approximately 0.015 mm^3 per gland; the skin contains over 2 million eccrine glands. Interestingly enough, the highest concentration of eccrine glands occurs in the palms and soles, areas that do not participate in thermal sweating. The chest and abdominal areas contain twice as many glands as the back. Although many lower species possess eccrine glands, only man has widespread distribution over the body surface.

Activity of the Eccrine Glands

The activity of individual glands varies with location and with intensity of heat stress. At low sweat rates, not all glands function; some do not secrete at all, others secrete intermittently, and still others may secrete continuously. With increasing heat stress, total sweat production increases, partly because some of the hitherto inactive glands are recruited, but mainly because the output of individual glands is increased. Some skin areas sweat more profusely than others; approximately 50 per cent of thermal sweat comes from the trunk, 25 per cent from the lower limbs, and 25 per cent from the head and upper limbs. The relative intensity is greatest on the head and trunk. The secretory pressure of sweat is high, 250 mm Hg.

Nerve Supply

Although the sweat glands receive only sympathetic innervation, the chemical mediator for thermal glands is parasympathetic (acetylcholine). However, the eccrine glands on the palms and soles conform to the usual sympathetic pattern in being adrenergic. Adrenergic sweating is a continuous process, and it is markedly and almost instantaneously increased with strong emotions or mental effort.

Blood Supply

Each gland receives its blood supply from an arterial twig, which branches into a large number of capillaries, the latter forming a basketlike network around the gland and duct. Kuno has calculated that, if all the sweat capillaries were laid end to end, their total length would exceed 50 miles [12]. Therefore, the vascularity of these glands appears sufficient to deliver large amounts of blood to support both sweat production and the glands' own metabolism.

Adequate Stimulus for Sweating

There is probably no single adequate stimulus for thermal sweating. Skin temperature is certainly an important stimulus, operating probably by way of warmth receptors; so also is core temperature, which acts by way of direct temperature effects on the hypothalamus. It can be shown that these factors may operate independently of each other. Thus a resting man begins to sweat when air temperature reaches 89°F. In this situation there is an increased skin temperature, but no elevation of core temperature. As the skin becomes warmer, sweat production becomes more profuse. On the other hand, sweating during exercise is due to an increased core temperature, often with no change or with even a drop in skin temperature. The stimulus for sweating when both exercise and heat are acting is probably a combination of skin afferents and blood temperature. Unlike emotional (adrenergic) sweat, thermal sweat requires a latent period for its appearance; this period may range from 2 to 8 min after heat exposure. Consequently, Adolph concluded that sweating is not a simple nervous reflex but requires a certain elevation of body-heat content for its onset [1].

The sensitivity of the sweat glands in terms of threshold and rate of excretion is variable, depending in part on what Rothman calls the "conditioning" of the glands [14]. This is exemplified by a seasonal variation in sensitivity. In summer, sweat is more readily induced, with a smaller latent period and at lower skin temperatures, than in winter. Even short, acute bouts of sweating appear to "condition" the sweat glands to a more sensitive response to subsequent heat exposures. Little is known regarding the mechanisms that control the sensitivity of the sweat glands. Other factors affecting the sensitivity of sweat glands are exercise, sleep, and cold. Sweating with exercise starts at a lower skin temperature, at a higher rectal temperature, and with a longer latent period than when the body is at rest. During sleep, sweating occurs at lower skin and rectal temperatures than during the awake state. Cold exposure results in complete inactivity of the sweating mechanism, with a subsequent increase in threshold and latent period on rewarming. Finally, intense heat stress considerably reduces the latent period for the onset of sweating.

Composition of Sweat

Sweat is over 99 per cent water and is therefore the most dilute of the body's secretions. More than 40 different chemical entities are present in the 1 per cent or less of sweat solids. The most important of these are ions of sodium, chloride, and potassium. None of the sweat solutes serves

a useful function in thermoregulation; the physiological role of sweat appears to be solely that of providing water for evaporative cooling. The sweat glands, therefore, do not regulate the composition of the blood; this function is performed by the kidneys and lungs; indeed, sweat losses of salt place an additional burden on the kidneys. The presence of salt in the sweat is, however, not an unmitigated physiological disadvantage, even though the secretion of pure water would be theoretically more efficient. The physiological benefit stems from the fact that the body fluids that are left in the body during sweating have homeostatic requirements for tonicity and composition, especially as regards sodium and potassium. The loss of these ions in the sweat mitigates the strain with respect to these parameters; if pure water were excreted, dangerously high concentrations would be reached sooner.

SHIVERING

Organization of Skeletal Muscle

Of the three organs directly concerned with thermoregulation, the cutaneous blood vessels and sweat glands differ from the skeletal muscles in two characteristics: (1) both skin organs act by modifying physical avenues of heat loss, and (2) they both are innervated by the sympathetic nervous system. The skeletal muscles, on the other hand, (1) serve thermoregulation by adding metabolic heat through shivering when depletion of body-heat stores is threatened and (2) are innervated not by autonomic but by somatic motor nerves. These muscles serve three major functions: voluntary movements, reflex maintenance of posture (particularly against gravity), and shivering.

Physiology of Shivering

Shivering differs from the other two skeletal-muscle functions in that it is uncoordinated; it may involve small groups of fibers or an entire muscle; it is not phased with antagonistic muscles, i.e., flexors and extensors, and, as a result, gross movement is minimized. The sole contribution of shivering to the body economy is heat production. Current concepts of the physiology are as follows: Control of shivering resides in the posterior hypothalamus. The stimulus for shivering may arise either from cold receptors in the skin or from a drop in core temperature. Either stimulus can act alone to produce shivering; conversely, shivering can occur with an elevated body temperature or in the presence of a

rising skin temperature. Thus a nude, inactive man will shiver within the first half hour of exposure to 50°F; his rectal temperature may be 1°F higher than normal, but his skin temperature will have decreased. In this situation, the stimulus for shivering is from the periphery, i.e., the result of skin cooling. Often an already chilled man who is not shivering will begin to shiver if he performs exercise; during the exercise his skin will rewarm, and his rectal temperature will decrease slightly. In this situation the prepotent stimulus is probably central, by way of a direct temperature effect on the brain. There is much controversy, similar to that about sweating, as to whether the normal adequate stimulus for shivering is peripheral or central. It is likely that both stimuli are operative, depending on rate of heat loss, muscular activity, and number of cold receptors affected; the "thermal state of the body," an admittedly diffuse expression, is an important factor in shivering. Whatever the adequate stimulus, it is transduced by the heat-conservation center into shivering, by way of efferent pathways that ultimately impinge on the muscles. These pathways are not well known. From the recent work of Hemingway [9] they have been traced from the shivering center, through the midbrain, and to the medulla, but no clear-cut tract has been demonstrated. The course from the medulla down the spinal cord is almost a complete mystery; it apparently is not by way of the pyramidal tract, which is the great motor pathway that subserves voluntary muscular activity.

Once shivering is present, what stops it? The obvious answer is the decrease or cessation of the afferent stimuli that initiated it. This, however, is an oversimplification. The cessation of shivering is probably the result of an integration of decreased cold afferents and increased stimulation of the anterior hypothalamus. There appear to be reciprocal connections between the two thermoregulatory centers such that stimulation of one suppresses the activity of the other. Thus stimulation of the heat-dissipation center can inhibit even violent shivering in a thoroughly chilled animal. If a chilled, shivering man is removed to a warm room, shivering rapidly ceases, even though he is still in heat debt and his skin is cold; this may be mediated by way of stimulation of the heat-dissipation center and accompanying inhibition of the shivering center.

Some Characteristics of Shivering

Latent Period. Like sweating, shivering requires a latent period; this period is variable, depending in large part on the rate of heat loss. For example, nude men will shiver within minutes of entering a room at 40°F with a wind of 10 mph; on the other hand, they may withstand 1 hr or longer at 60°F without shivering, even though they are in heat

debt and their skin is cool. With rapid cooling, it is possible that the cold receptors display the overshoot already described and that a sufficient number of receptors are cooled below the point of adaptation. Normally, an individual has already incurred a body-heat debt when shivering begins. This debt may range as high as 100 kcal.

Efficiency. Although violent shivering can add metabolic heat to the extent of 400 kcal/hr, it is so graded that it cannot completely repay the body-heat deficit, and therefore it serves only to retard further net heat loss. Voluntary activity has a far greater capacity for increasing heat production and is therefore much more "efficient" than shivering; this capacity is more than five times that of shivering.

Inhibiting Factors. Voluntary exercise inhibits shivering; the mechanisms are not clearly understood. Breathing pure oxygen or high concentrations of carbon dioxide also depresses shivering, as does deep anesthesia and hypothermia. Shivering is also affected by conscious thoughts; one can inhibit mild shivering by will power, or shivering can be exaggerated by anticipatory apprehension of cold discomfort.

FAILURE OF THERMOREGULATION

Some insight into the importance of thermoregulation can be gained if one examines the effects of excessively heating or cooling the body. Death will almost certainly occur if the internal body temperature exceeds 110°F, and it may occur if core temperature reaches below 78°F. Present techniques for inducing hypothermia have, however, greatly extended the lower temperature for survival under controlled conditions.

What are the causes of death in these situations, and what irreversible damage takes place in the tissues? In heat stroke, the sequence of events is as follows: First there is exposure to excessive heat load. This results in an unduly high core temperature, which is associated with a decrease in sweat production. As sweating decreases, the body temperature rises still further and there ensues a dangerous chain of events, in which sweating is more and more decreased until it ceases entirely. The absence of sweat is a cardinal danger sign, for the body, entirely deprived of its major defense against heat, stores more and more heat until cells in the vasomotor centers of the brain stem are damaged. Until this occurs, the vascular system maintains optimal distribution of blood flow; i.e., the vessels of the viscera are constricted, and those of the skin are dilated. Now, however, deprived of central nervous controls, the visceral vessels dilate, and the individual enters a state of shock, the result of excessive pooling of blood in the viscera. As a result, there is inadequate blood flow, not only to the skin, but also to the brain. Heat piles up at an ex-

plosive rate, and the brain suffers from both high temperature and lack of oxygen, a fatal parlay. Post-mortem examinations reveal brain damage in the regions of the cerebral cortex, cerebellum, and upper brain stem, the latter containing important cardiovascular and respiratory control centers. Surprisingly little or no damage is found in the hypothalamus itself, nor are the sweat glands damaged. Death from overheating is due to direct damage by heat to important brain centers. The damage is probably the result of an irreversible change in the brain proteins.

Death from cold (hypothermia) is interestingly different from heat stroke. The core temperature can be reduced to approximately 78°F with no ill effects. At this temperature the individual is unconscious and does not shiver; all metabolic activity is considerably reduced, and there is an accompanying reduction in requirements for oxygen. For example, the brain may receive only 25 per cent of the oxygen normally required and yet not be damaged; at normal body temperatures an oxygen reduction of this magnitude would lead to severe brain damage or even to death from anoxia. Coronary blood flow is similarly reduced to levels that would result in cardiac damage under normal conditions but that adequately meet the lower requirements during hypothermia. Death finally occurs at core temperatures ranging from 75°F down to 60°F. The time and cause of death are not predictable, but the greatest dangers are to the heart muscle (particularly the ventricles) and to respiration. The heart may simply stop beating in a relaxed state, or it may go into ventricular fibrillation; this may occur during the cooling period or during a rewarming phase. Why the heart or respiration fails is not clearly known; however, the failure is due not to destruction of tissue but rather to impairment of cellular function. This may be the result of disturbances in ionic and acid-base balance that in turn affect enzymatic processes and irritability, or it may be due to slowing of metabolism to the point where integrated chemical reactions cease.

In summary, death from hyperthermia is the result of irreversible damage to body cells. Death from cold is due to functional impairment of cells without discernible structural damage. It is of interest that the central nervous system, which is very sensitive to heat, is among the most resistant of all organs to functional impairment from hypothermia. The heart, in contrast, is one of the most vulnerable of the core organs to functional disruption by cold.

UNIFYING CONCEPT OF THERMOREGULATION

The most obvious function of thermoregulation is to maintain an optimal range of temperatures (97 to 105°F) in some area of the brain

and thereby to provide an optimal temperature for all cellular function. To consider its role solely in such terms would represent a limited viewpoint indeed. For homeothermy, like all homeostatic mechanisms, is in its broadest sense a way of life; i.e., its ultimate role is the preservation, not of a temperature, but of an optimal thermal environment for the multiplicity of integrated chemical reactions that constitute life.

At first glance it appears that man is better able to survive cold than heat stress. It is true that the body normally operates at core temperatures that are much closer to upper than to lower lethal limits. If, however, man is regarded as a total organism that must function optimally in order to survive, cold appears to be as much a threat to survival as heat in terms of (1) the degree of cooling that he can tolerate and (2) the purely physiological defenses at his disposal.

Regarding tolerances, man becomes incapable of conscious, lifesaving decisions when deep-body temperature is lowered by only 7°F and is therefore unprepared for "fight or flight." Furthermore, long before this point is reached, he has become subjectively so uncomfortable that during cooling he must devote 100 per cent of his effort to mere survival. In contrast, well-motivated men in good physical condition, e.g., athletes and soldiers, can function in the heat, without conscious awareness of the need for survival measures, right up to the point of collapse and death from heat stroke. Regarding defenses, the shivering mechanism can add to the body a maximum of only about 350 to 400 kcal/hr, whereas sweat production and evaporation can dissipate four to five times this amount of heat. Man's physiological defenses, therefore, are greater against heat than against cold. However, heat poses a greater threat to life because (1) metabolic heat is continually produced, (2) normal body temperature is closer to upper than to lower (lethal) limits, and (3) excessive motivation, e.g., athletics and survival action, can cause a man to ignore physiological warning signals of excess heat stress. Cold, on the other hand, poses an immediate threat, not to life, but rather to normal function, for (1) frostbite occurs readily, (2) hand dexterity is impaired, and (3) shivering can become so violent that coordinated gross body motions are impossible.

The key terms of a unifying concept of thermoregulation are not heat, cold, and temperature but rather life, death, and survival. Protection against hyperthermia is protection against death itself. Protection against hypothermia is protection of normal function that is necessary for coordiated lifesaving decisions of "fight or flight." When the core temperature drops sufficiently to abolish consciousness, the body "lets go" of thermoregulation long before death is imminent; in hyperthermia, thermoregulation is so tenaciously maintained as to permit the organism to fight literally to the death. The survival advantage of the latter is obvious.

BIBLIOGRAPHY

1. Adolph, E. F.: "Physiology of Man in the Desert," Interscience Publishers, Inc., New York, 1947.

2. Bard, P.: "Medical Physiology," 10th ed., chap. 48, The C. V. Mosby Company, St. Louis, 1956.

3. Behnke, A. R., Jr., B. G. Feen, and W. C. Welham: The Specific Gravity of Healthy Men: Body Weight ÷ Volume as an Index of Obesity, *J. Am. Med. Assoc.*, vol. 118, pp. 495–501, 1942.

4. Benedict, F. G., and F. B. Talbot: *Carnegie Inst. Washington Publ.* 503, 1938.

5. Bernard, C.: "Leçons sur la physiologie et la pathologie du système nerveux" ("Leçons d'ouverture," vol. 1), Paris, 1858.

6. Cannon, W. B.: "The Wisdom of the Body," 2d ed., W. W. Norton & Company, Inc., New York, 1939.

7. Carlson, L. D.: Man in Cold Environment, Arctic Aeromedical Laboratory, Alaskan Air Command, Fairbanks, Alaska, 1954.

8. DuBois, E. F.: "Fever and the Regulation of Body Temperature," Charles C Thomas, Publisher, Springfield, Ill., 1948.

9. Hemingway, A., P. Forgrave, and L. Birzis: Shivering Suppression by Hypo-thalamic Stimulation, *J. Neurophysiol.*, vol. 17, pp. 375–386, 1954.

10. Hendler, E., and J. D. Hardy: Temporal Aspects of Temperature Sensation, *ASME Publ.* 58-A-220, 1958.

11. Hensel, H., and Y. Zotterman: The Response of the Cold Receptors to Constant Cooling, *Acta Physiol. Scand.*, vol. 22, pp. 96–105, 1951.

12. Kuno, Y.: "Human Perspiration," Charles C Thomas, Publisher, Springfield, Ill., 1956.

13. Newburgh, L. H. (ed.): "Physiology of Heat Regulation and the Science of Clothing," W. B. Saunders Company, Philadelphia, 1949.

14. Rothman, S.: "Physiology and Biochemistry of Skin," University of Chicago Press, Chicago, 1954.

Chapter 3

HUMAN PERFORMANCE AND THE WORK-REST SCHEDULE

Oscar S. Adams and W. Dean Chiles†*

A SIGNIFICANT FEATURE OF MANY FUTURE MAN-MACHINE SYSTEMS IS THE extent to which they will depend on efficient utilization of crew members. This is particularly true for advanced systems in which (1) weight is a critical factor, (2) projected mission durations exceed 30 hr, and (3) there is a requirement for around-the-clock performance of one or more tasks. Since severe weight penalties and logistical problems might be incurred by carrying alternate crews and their associated equipment, it will be necessary to obtain maximum utilization of a minimum number of personnel. For example, systems such as long-range (nuclear-powered) aerodynamic vehicles, earth-orbiting satellites, and space vehicles will probably be operated under circumstances in which personnel cannot be readily added or exchanged.

One way to increase the crew-utilization factor is to schedule the individual operator's phases of work and rest in such a way that the proportion of on-duty time is the maximum that can be obtained with a high probability of efficient job performance. For the purpose of this discussion it is assumed that in the system there are certain limiting factors requiring that the total amount of time spent at work should equal and preferably exceed the total time spent off duty. Achieving this involves consideration of three aspects of the work-rest schedule: the length of the individual work period, the length of the individual rest period, and the ratio of the amount of time spent at work during a 24-hr period to the total amount of time spent at rest, sleep, or nonwork activities. To illustrate, a 1:1 ratio of work and rest could involve 2-hr shifts, or it

* Human Factors Research Department, Operations Research Division, Lockheed-Georgia Company, Marietta, Ga.

† Training Research Branch, Behavioral Sciences Laboratory, Aerospace Medical Laboratory, U.S. Air Force, Wright-Patterson Air Force Base, Dayton, Ohio.

The research reported in this chapter was supported by the Behavioral Sciences Laboratory, Aerospace Medical Laboratory, Aeronautical Systems Division, U.S. Air Force Systems Command, under Contract AF 33(616)-6050 with the Lockheed-Georgia Company, Marietta, Ga.

could involve 12-hr shifts; a 2:1 ratio could consist of 4 hr work and 2 hr rest, or it could consist of 8 hr work and 4 hr rest.

Man's ability to adapt himself to a specific work-rest schedule may be regarded as being a function of two arbitrary classes of variables. Class I variables consist of those conditions which affect the general psychological and physiological predisposition of the individual to adjust to a rigorous work routine, and they include factors such as level of motivation, physical health, emotional adjustment, and attitude. For the purpose of this discussion it is assumed that these factors can be optimized through the use of appropriate selection and training techniques. Balke [6], for instance, has shown that with proper training and motivation man can adapt to almost superhuman requirements of a biological nature.

Class II variables consist of those behavioral and physiological parameters which, in effect, limit the extent of individual adjustment even though Class I conditions are optimum. In some cases they might be identified as *physiological limits,* and in others they might be described more appropriately by a concept such as *channel capacity.* There are at least four important factors that are characteristic of this class of variables: (1) the amount of sleep required by man, (2) the nature and duration of the work task, (3) the adaptability of man's natural diurnal rhythms, and (4) the level of proficiency demanded during prolonged performance. It is believed that a lack of compatability between any one of these factors and a specific work-rest schedule will lead to a decrement in performance and thus to a suboptimal level of proficiency.

Unfortunately, the Class II variables might well be expected to interact; i.e., the amount of sleep required to perform at a satisfactory level on one type of task is very likely to be different from that required for other types. In searching the literature for information on this problem one finds that many of the studies that appear to be relevant actually turn out to be of only limited usefulness, for one or more of four main reasons: (1) the situations studied did not involve a high crew-utilization factor; (2) the tasks were not representative of those found in advanced man-machine systems; (3) the data were not collected in quantitative form; (4) no control was exercised over the off-duty activities of the subjects. In this respect, one finds that the industrial literature, most studies of crew performance in both surface and underwater craft, and reports concerning isolated radar sites are not directly applicable to the work-rest problem as herein conceived. In a recent review of the literature, Ray, Martin, and Alluisi [55] concluded that, while there are a large number of studies that provide clues as to the operation of these factors considered separately, there are very few that avoid the above limitations to any appreciable degree of satisfaction.

SLEEP REQUIREMENTS

Most of the studies pertaining to sleep requirements have dealt with the average duration of the sleep period under normal conditions or with the effects of prolonged wakefulness rather than the minimum amount of sleep required for maintaining efficient performance. Those investigations which have focused attention on established periods of sleep in adults have found them to range from approximately 6 to 9 hr. In addition to depending on differences within and between individuals, this "standard" sleep period may be influenced by factors such as age, time of going to bed, mode of awakening, and depth of sleep [38].

The results of sleep-deprivation studies essentially agree that the degree of performance decrement during prolonged hours of wakefulness is a function of the length of the performance-testing period and the type of task involved. In general, no pronounced decrements are observed if the tasks are simple and the period of performance is short. After reviewing the results of his own extensive studies as well as those of other investigators, Kleitman [38] concluded that the mental and muscular performance of sleep-deprived subjects is normal during short periods of measurement but subnormal for sustained effort.

A number of studies conducted during the past 20 years have corroborated the findings of earlier investigations. Edwards [14] found that subjects after a loss of 100 hr of sleep showed inconsistent performance or no change in short tests of auditory-reaction time, hand steadiness, and hand grip. Performance on the American Council on Education Psychological Examination, memory for nonsense syllables, and static ataxia did not begin to deteriorate until the last 30 hr of wakefulness. Further confirmation of these results was obtained by Bliss, Clark, and West [7], who were unable to find any significant differences between scores obtained on a nonsense-syllable association test and on a mutilated-word test presented daily during a period of 72 hr of wakefulness. They concluded that sleep-deprived subjects have the capacity to mobilize their intellectual facilities when confronted with formal tests of relatively short duration.

The effect of extending the period of performance testing was demonstrated very clearly in a 112-hr experiment conducted by Tyler [58] with 275 servicemen. Beginning about the sixtieth hour a noticeable difference in performance occurred when the duration of a multiple-choice reaction-time test increased from 2 to 10 min. Likewise, the loss of sleep had little effect on marksmanship scores obtained with 10 to 15 rounds of ammunition but did lower the scores based on 68 rounds of ammunition.

That performance of complex tasks is affected more than performance

of relatively simple disjunctive tasks has been demonstrated by Clark et al. [13]. Every 6 hr during a 50-hr vigil they presented a test period consisting of 15 min of stereorange settings with the Mark II Navy trainer and approximately 1 hr of testing with an alertness apparatus. They observed that the level of performance on the complex task (alertness apparatus) decreased over the 50-hr period, whereas performance on the simple task (stereoranging) did not change significantly. More recently, Chiles [11] found that, when compared with a control group, subjects who were deprived of sleep for a period of 30 hr were unable to show the same degree of improvement in the performance of a complex mental task that they had not previously mastered.

Ax et al. [5] have suggested that intrinsic task interest is a dimension that must be considered in interpreting the effects of sleep deprivation on performance. Following 24 hr of wakefulness, subjects were presented a battery of 18 psychomotor, attention, and performance tests for a period of 8 hr. Performance of the low-motivating tasks decreased markedly, whereas high-motivating tasks were performed without decrement. The authors interpreted their findings as indicating that moderate sleep deprivation lowers the internal-drive state and that this condition is temporarily compensated for by increased external motivation.

One of the most consistent observations made during the course of studying the effects of sleep deprivation concerns the clinical behavior of the subjects. After 24 to 30 hr of wakefulness subjects are very difficult to keep awake unless they are forced to be physically active or are continually prodded by the experimenter. Reports range from accounts of irritability, inattention, and uncooperativeness to incidences of hallucinations, bizarre behavior, and episodic rage. Edwards [14] found that his subjects were able to take the American Council on Education Psychological Examination only through greatly increased effort. Bliss, Clark, and West [7] reported that, although Rorschach profiles were normal both before and after 72 hr of wakefulness, feelings of depersonalization, illusions, hallucinations, disturbances of time perception, and auditory changes were common.

Bulban [9] has reported Hauty's description of the experiences of four nonpilots and four pilots who were exposed individually to a simulated space-flight environment for 30 hr in the U.S. Air Force School of Aviation Medicine's one-man sealed chamber. The conditions included the subject's performing a battery of tasks continuously while confined to the 50-ft^3 chamber, which was maintained at an atmospheric pressure of approximately 380 mm Hg (18,000 ft altitude). During and after the "flight," all four of the nonpilots reported hallucinations (e.g., gremlins, melting instruments, and a burning television monitor), and two of the pilots reported less dramatic experiences (vertigo and visual illusions).

It is quite possible that this difference between the two groups can be explained by the nature of the pretest instructions. The author states that the nonpilot subjects were acquainted beforehand with the probable occurrence and nature of the bizarre experiences that they might encounter. The pilot subjects, on the other hand, were not given any prior instructions as to what they might experience in the chamber.

Behavioral and subjective alterations are probably most prominent during extremely long periods of wakefulness. Katz and Landis [36] have described a young man who, convinced that sleep was only a habit that could be broken, volunteered as a subject to test the hypothesis. At the end of 231 hr he became so disturbed that it was necessary to terminate the experiment. A similar account is given by Luby et al. [48], who studied a subject participating in a 220-hr "wakethon." Deficits in visual-motor performance occurred cyclically across days and gradually deteriorated to a point of virtual untestability on the ninth day.

Since there are no systematic data specifically related to the problem, the question of minimum sleep requirements for extended periods of time cannot be answered unequivocally. Laird and Wheeler [44] found that mental multiplication was unaffected by shortening the normal sleep period by 2 hr nightly for a week. Likewise, Freeman [19] could observe no consistent trend in discrimination reaction, manual pursuit, and memory span when he and his wife systematically varied the duration of their sleep to include seven periods each of 4, 6, 8, and 10 hr for 28 consecutive days. Husband [30] studied the effects of a divided-sleep schedule in which a single subject was permitted to sleep for 3 hr, was kept awake for 3 hr, slept for 3 hr, and then stayed awake for 15 hr over a period of 1 month. He found no consistent differences between performance scores obtained under this schedule and those obtained during a 1-month period in which the subject slept 8 consecutive hours nightly.

After an extensive program of physical preconditioning, Balke [6] and another subject spent 10 days in a small low-pressure chamber that was kept at barometric pressures equivalent to 14,000 to 18,000 ft. High temperatures, humidity, and other environmental conditions caused restlessness, headaches, and nausea to such an extent that only 3 to 4 hr of sleep was obtained by each subject during each 24-hr period. It was found, however, that the number of addition problems completed in a 1-hr arithmetic test presented daily increased steadily during the confinement period; error and correction scores were found to increase steadily also [21]. Furthermore, both subjects maintained a state of alertness sufficient to make routine checks of equipment and to perform accurate calculations and instrument manipulations.

In a study by Adams and Chiles [2], 4 groups of 4 subjects each were assigned to one of 4 work-rest schedules—2 hr on and 2 hr off, 4 hr

on and 4 hr off, 6 on and 6 off, and 8 on and 8 off— for a period of 96 hr. At the end of testing, 13 of the subjects reported that they had got enough sleep, which was estimated to average a little more than 8 hr out of 24. The levels of performance on tasks involving arithmetic computation, pattern discrimination, monitoring, and vigilance continued to improve over the 4-day period for all groups, and no significant between-group differences were obtained. A subsequent study by the same investigators employed schedules of 4 hr on and 2 hr off and 6 hr on and 2 hr off, with 10 subjects being assigned to each schedule for 96 hr. It was estimated that subjects on the 4-2 schedule obtained, on the average, 5½ hr of sleep, whereas those on the 6-2 schedule were able to get only about 4 hr of sleep during a 24-hr period. Sixty per cent of the subjects in both groups indicated that this was enough sleep. When asked whether they would volunteer to undergo the same experience again, 80 per cent of the subjects on the 4-2 schedule and 50 per cent of those on the 6-2 schedule replied affirmatively. Performance improved on some tasks and decreased on others, and no consistent differences were observed between the two groups.

Data from these and other studies, along with everyday experience, suggest that 6 hr of sleep per day are adequate for most people. If sleep is to be divided among two or more periods, certain rational considerations suggest that 2 hr is probably the shortest acceptable duration for a subperiod. This results from the fact that the amount of time required to fall asleep (approximately 20 to 30 min) would consume an unreasonable proportion of a rest period that is of less than 2 hr duration.

DUTY-PERIOD DURATION

The primary factor to be considered in selecting an appropriate duration for the duty period is the nature of the activity required of the operator. Some tasks involve relatively passive performance on the part of the individual in that he may sit for periods of several minutes while waiting for an event that happens at a very low average rate; this sort of task is exemplified in radar watchkeeping. On the other hand, some tasks involve active participation of the operator by requiring that he perform some actions more or less continuously. Manual control of an aircraft and many industrial machine operations are examples of this type of task. Most tasks, of course, fall somewhere between these two extremes.

Laboratory studies, such as those of Mackworth [50] and Jerison and Wallis [33], on the problem of vigilance provide some insight into the effect of the length of the duty period on the performance of passive

tasks. When subjects are presented with simulated radar-watchkeeping tasks or some variation of the British clock test, decrements in the detection of infrequent (30 to 40 per hour) signals occur as early as 30 min after the beginning of the test period. Lindsley et al. [47], studying the performance of radar operators who were on duty 4 hr/day, 6 days a week, for 17 days, found both within- and between-day deterioration in efficiency. After the first few days, this impairment reached statistical significance after the first 40 min of morning operations and after the second hour of afternoon operations. Although Webb and Wherry [59] found an increase in the latency of responding to auditory signals presented late during a 9-hr session, they did not observe any predictable change between five sessions repeated on successive days.

Elliott [15] has recently questioned the applicability of laboratory vigilance studies to practical watchkeeping situations. He states that results obtained from studies of real military watchkeeping fail to confirm some of the published data in that "fatigue" effects are commonly absent during continuous 2-hr watches. He attempts to explain some of the discrepancies and suggests several experimental designs that might be studied in order to resolve the differences. For a review of the major study findings and principal theories of vigilance, the reader is referred to a recent report by Frankmann and Adams [16].

These decrements can usually be prevented, or original levels regained, by introducing distractions, time sharing with other tasks, or changes that tend to relieve monotony. For example, Gaito et al. [20] found decrements in the performance of one passive task presented by itself but not in a second one that was combined with an active task. Gorham, Orr, and Trittipoe [23] presented two vigilance tasks continuously in combination with each one of three psychomotor tasks for a period of 24 hr. The performance curves for their two subjects show that all tasks resisted impairment until about the sixth to tenth hours. The results of the Adams and Chiles study [2] were similarly suggestive. Subjects who followed a 2-2 or 4-4 schedule did better on the monitoring tasks than did those subjects who followed an 8-8 schedule. Furthermore, from responses to a posttest questionnaire, it was found that the 2-2 and 4-4 subjects seemed to make a better overall adjustment to what was a boring situation.

As for active tasks, it has been amply demonstrated in the industrial situation that, in those jobs in which the worker has control over his rate of activity, the production rate per hour will increase if the length of the workday is decreased from 12 or 10 hr to 8 hr/day [22]. In addition, very valuable related gains are realized through decreases in accident rates and absenteeism. An observation commonly made in such situations is that man works at a near-maximum rate for a

period and then takes either an unofficial or official "rest break," after which he usually resumes his original rate of output. Because of these breaks, the period of continuous duty for most industrial jobs is usually about 2 hr and, because of lunch breaks, is seldom longer than 4 hr.

Many of the results from laboratory studies are sometimes difficult to interpret because of the lack of comparability of tasks and task conditions, the differences in motivational levels of the subjects, and the differences in the number of replications of the performance periods. Huxtable, White, and McCartor [31], repeating an earlier study by Arai [4], have published performance curves resulting from 12 hr of continuous mental multiplication of 4-digit numbers on each of 4 successive days. A within-day effect consisting of an increase in time required to solve problems was reported in both studies, but it was less prominent in the Huxtable study (see Figure 3-1). Also noted was a

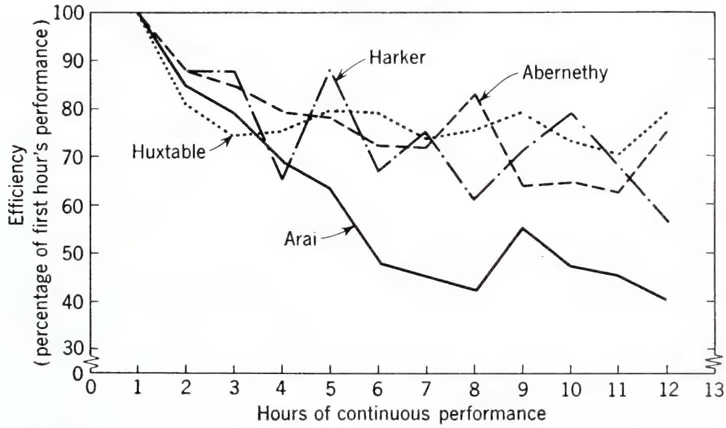


FIG. 3-1 *Efficiency (percentage of first hour's performance) attained by Arai [4] and Huxtable et al. [31] during each of 12 consecutive hours spent in mental multiplication of 4-digit numbers. Each point represents the average of a single subject's performance over 4 days as measured by the average time per successful operation. (After J. T. Ray, O. E. Martin, and E. A. Alluisi, Human Performance as a Function of the Work-Rest Cycle: A Review of Selected Studies, Natl. Acad. Sci.–Natl. Research Council Publ. 882, 1961. By permission of the publishers.)*

between-day effect associated with an increase in the speed of problem solving but a decrease in the number of problems solved correctly.

Payne and Hauty [29, 53, 54] have made an extensive study of the effects of motivation, drugs, and knowledge of results on the performance of a fairly active task, the U.S. Air Force SAM Multidimensional Pursuit Test. Typically, the subjects' practice curve leveled off about 30 min after the beginning of a 50-min practice session. This was followed by

a 10-min rest period, during which subjects were assigned to different experimental treatments. Except for those conditions involving the administration of analeptic drugs, negative performance trends usually occurred during an ensuing 4-hr test period. Far less pronounced effects were obtained in a later study [28], in which the same task was performed for a period of 30 hr. In fact, the investigators present data that show that the initial level of performance was maintained for as long as 16 hr. Why there should be this wide discrepancy between the results obtained with short- vs. long-term duty periods is not easily explained unless it can be attributed to differences in subject motivation. In another paper, Hauty [27] offers a clue when he points out that a general "fatigue" curve often obscures wide variations among individuals (see Figure 3-2).

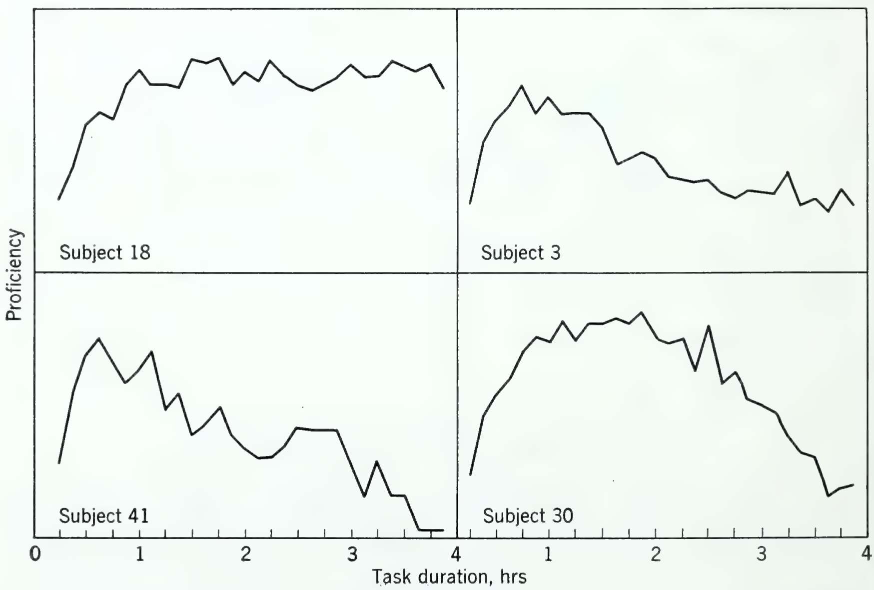


FIG. 3-2 *Differences in trend of individual proficiency curves of four subjects performing an active task for 4 consecutive hours. (After G. T. Hauty, Human Performance in the Space Travel Environment, Air Univ. Quart. Rev., vol. 10, pp. 89–107, 1958. By permission of the publishers.)*

Except for three short (10-min) food-relief breaks, subjects in the Gorham, Orr, and Trittipoe study [23] were required to operate alternately for 24 hr one of three psychomotor tasks selected to measure eye-limb coordination (SAM Complex Coordination Test), rate-of-closure judgment (SAM Motor Judgment Test), and the ability to select and manipulate controls (SAM Direction Control Test). In addition, two

vigilance tasks and one problem-solving task were programmed concurrently, but on a time-shared basis, with the psychomotor tasks. Performance did not begin to deteriorate until about the sixth to tenth hours, confirming somewhat the results of earlier 8-hr tests in which very little decrement occurred.

Of direct relevance to the problem of work-task duration are studies that have measured flying proficiency during long instrument flights [32, 49], navigator activity during extended arctic missions [12], and radio-operator activity during long reconnaissance flights [35]. These studies suggest that man may lower his performance standards toward the end of a 12- to 17-hr flight. Further evidence of this effect is reflected in poorer vigilance and problem-solving test scores obtained after long flights as compared with those recorded prior to flight or following a controlled rest period [17, 18, 60]. In addition, investigations of truck drivers after 10 hr of driving found decreases in efficiency on tests of reaction coordination, reaction time, manual steadiness, and driving vigilance [34].

On the basis of these and other data from similar studies, an interval of 4 hr is recommended as the maximum duration for a duty period when (1) a passive task is combined with one or more active tasks, (2) the work load is not too great, and (3) a high level of performance must be maintained continuously. When a passive task occurs by itself, a 2-hr duration is probably the maximum advisable. The duty period may be routinely extended to 8 to 10 hr if the major tasks call for active participation, if there is considerable variety in the tasks, and if passive tasks have very readily detectable signals to which the operator must respond. Duty periods that require relatively continuous performance for longer than 10 hr are likely to require that crew members exert increasing effort in order to avoid lowering of performance standards. In such cases, a lapse in effort would undoubtedly decrease the probability of mission success.

DIURNAL RHYTHMS

The fact that certain rhythmic or cyclic changes in man's physiological processes are associated with alternate phases of rest and activity or sleep and wakefulness has been thoroughly documented. Various terms have been used to designate this phenomenon. To emphasize the correlation with the 24-hr day, some writers use the terms "diurnal rhythm" (or "cycle") and "circadian periodicity" [24]. Others [57] refer to it as the "physiological day-night cycle," since the period of wakefulness and activity normally occurs during daylight hours and the period of rest

and sleep is usually confined to night hours. This rhythmic feature of the biological processes is manifested by changes in respiration rate, oxygen consumption, gastrointestinal activity, heart rate, body temperature, blood pressure, and many other metabolic functions. It has been observed, for instance, that there are regular diurnal variations in blood-cholesterol level that are independent of food intake [51].

Kleitman [37, 38], who has conducted the most extensive study of this phenomenon, has suggested that body temperature is probably the most sensitive index of measurement. This is illustrated in Figure 3-3,

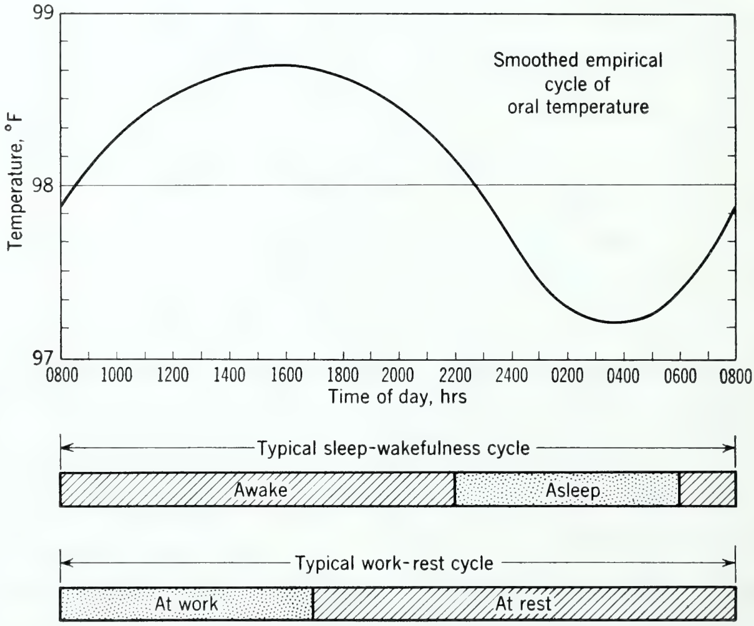


FIG. 3-3 *Diurnal variation in man's oral temperature. Smoothed curve based on empirical data obtained by Kleitman* [39]. (*After J. T. Ray, O. E. Martin, and E. A. Alluisi, Human Performance as a Function of the Work-Rest Cycle: A Review of Selected Studies, Natl. Acad. Sci.–Natl. Research Council Publ. 882, 1961. By permission of the publishers.*)

in which the mean hourly readings of oral temperature have been plotted for a period of 24 hr. These data were obtained from one man over the course of 9 full days. When smoothed, the curve follows a monophasic cycle, with a maximum temperature occurring during the wakefulness period and the minimum temperature occurring during the hours devoted to sleep. It is fairly typical of the physiological diurnal cycle found in adult man and most higher animals. In this connection it is interesting to note that the human infant follows a polyphasic cycle. Through a

process of acculturization this gradually changes so that, by the age of ten, monophasic sleep habits are established rather firmly.

Attempts to eliminate the diurnal rhythm by exposing the organism to long periods of wakefulness have proved unsuccessful. Kleitman and Schreider [42] tested the ability of subjects, kept awake overnight, to maintain visual fixation on a point for 5 min at 2-hr intervals. The incidence of diplopia resulting from an ocular imbalance increased with time, becoming most prominent during the early hours of the morning. Later in the day, however, the subjects became more wide-awake, and good binocular fixation was gradually restored. Murray, Williams, and Lubin [52] found a persisting diurnal variation as well as a decrease in the daily level of body temperature in subjects deprived of sleep for 98 hr.

There is some evidence that the time-associated physiological changes can be adapted, within limits, to periods of less than and greater than 24 hr. Although he observed wide individual differences, Kleitman [38, 41] reports that it is possible to adapt body temperature and, to a greater extent, heart-rate rhythms to an 18-hr and to a 28-hr cycle. This conclusion is based on observations made over a 9-week period of living in Norway during the polar summer and also those made during 2 months of living in the Mammoth Cave in Kentucky. Brindley [8], on the other hand, was unable to find any evidence of complete physiological adaptation in eight men living on a 22-hr cycle for 8 weeks. The urine volume and pH and the sodium- and potassium-excretion rates of all subjects showed some signs of a persisting 24-hr rhythm with some superimposed 22-hr periodicity and random variation.

Lewis and Lobban [45, 46] have published data indicating that the various physiological rhythms are not equivalent with respect to their adaptability to a non-24-hr routine. They recorded the body temperature and the excretion of water, potassium, and chloride of one group of subjects living on a 21-hr day and of another group living on a 27-hr day for a period of 7 weeks in Spitzbergen. Body-temperature rhythms of 11 of the 12 subjects adapted almost immediately to the abnormal routine, but the excretory rhythms showed immediate adaptation in only 3 subjects. Although there was some evidence of improvement in adaptation of the excretory rhythms, it was never complete. Furthermore, small but statistically significant differences between the phases of the excretory rhythms were obtained, suggesting that more than one mechanism controls the physiological diurnal rhythms in man. As shown in Table 3-1, Lewis and Lobban have attempted to classify some of the diurnal rhythms with respect to their modifiability.

There is a great deal of evidence that indicates that performance efficiency has a diurnal variation closely resembling that found in phys-

TABLE 3-1 *Classification of Diurnal Rhythms According to Their Relative Responses to Life on Abnormal Time Routines*

Rhythm	Type of response to an abnormal routine
Heart rate Blood pressure	Entirely environmental
Phosphate excretion Body temperature	Largely environmental, though evidence of an inherent 24-hr rhythm in some subjects
Water excretion Chloride excretion Sodium excretion	Marked inherent 24-hr rhythm, which may be very pronounced in many subjects
Potassium excretion	Very marked 24-hr rhythm, usually predominating over any environmental influence

SOURCE: P. R. Lewis and M. C. Lobban, Dissociation of Diurnal Rhythms in Human Subjects Living on Abnormal Time Routines, *Quart. J. Exptl. Physiol.*, vol. 42, pp. 371–386, 1957.

iological processes. Kleitman [38] has demonstrated that the speed and accuracy curves for performing tasks such as mirror drawing, code transcription, multiplication, and card sorting show a well-marked diurnal rhythm that is very similar to that obtained for body temperature and heart rate. There is also a parallelism between the body-temperature curve and the performance of ataxia, hand-steadiness, and reaction-time tests.

Hauty and Payne [28] found that the proficiency curve for 24 hr of almost continuous performance on the SAM Multidimensional Pursuit Test was roughly equivalent to the body-temperature curve. Subjective ratings of sleepiness and fatigue by subjects deprived of sleep for 98 hr have shown an inverse correlation with measures of body temperature [52].

As seen from their long use by the armed services, man has little difficulty in adjusting to modified work-rest routines provided that they are systematically programmed around a 24-hr day. This is substantiated by Kleitman's World War II study [38] of the around-the-clock routine encountered on a fleet-type submarine at sea. The crew was divided into four sections, three of which followed a schedule of 4 hr on and 8 hr off and one of which followed a normal shore-type schedule of 8 on and 16 off. Kleitman reports that the 4-8 sections demonstrated double diurnal curves of both sleep-wakefulness and body temperature, whereas the 8-16 group demonstrated a curve similar to that shown in Figure 3-3.

In a subsequent study, Kleitman and Jackson [40] found that the diurnal body-temperature curve could be modified to conform to a variety of experimental activity schedules so long as they did not deviate too much from the usual routine of living. The "office-hour," or control, sched-

ule (0800 to 1200, lunch, 1300 to 1700) and the six experimental schedules never required more than 8 hr of work during any 16-hr period. In five of the experimental schedules the two or three work periods occurred at the same time each day. The sixth experimental routine varied the schedule of the work periods in such a way that they never occurred at the same time of the day during four successive days of a 12-day study. Under this routine of rotating activity, a 24-hr diurnal rhythm persisted in the curves for body temperature and color naming and to a lesser extent in those for reaction time and Link Trainer operation.

There is no question concerning man's propensity for developing physiological and behavioral rhythms that are in synchrony with a 24-hr day. Whether this is a genotypical characteristic of the human organism or whether it results entirely from a long process of conditioning is not known. Whatever the reason, it manifests itself as a strong and important influence on man's output efficiency. A conservative conclusion, therefore, is that scheduling of either single or multiple periods of work and rest should be systematically planned around a 24-hr day.

PROLONGED PERFORMANCE

Most of the research data that have been reviewed up to this point, in combination with everyday observations, support the conclusion that for periods of short duration (24 to 36 hr) man can, if necessary, mobilize his energy resources to offset the stressful effects of a heavy work program. In the final analysis, however, the feasibility of using any prescribed work-rest schedule depends on how well and how long it can be followed without seriously affecting the efficiency of the system.

Steinkamp et al. [56] have described an experiment in which five subjects individually followed a schedule of 4 hr on and 4 hr off while confined for a 7-day period to the U.S. Air Force School of Aviation Medicine one-man "space-cabin simulator." During the 4-hr duty period, subjects were required to perform a program of four tasks involving spatial discrimination, perceptual judgment, vigilance, and problem solving. Of the four subjects who completed the test, three maintained "their respective high initial levels of proficiency" throughout the period of confinement. The fourth subject showed a rather pronounced within-day variability and a gradual decrement in performance as the duration of the test increased. The investigators attributed this inability of the one subject to adjust to the unusual work-rest schedule to the fact that, unlike the other three subjects, he was a nonpilot and hence was unfamiliar with the "stresses encountered in the simulated flight."

Several physiological measures were also recorded during the 7

days of confinement. Oral temperatures, which were reported once every 8 hr by each subject just prior to the beginning of his work period, showed no systematic variations. Pulse and respiration rates for the periods of activity were higher than those recorded for the periods of rest, but there was no clear indication of a 24-hr periodicity.

Two studies of the effects of prolonged confinement have been reported by investigators at the Naval Air Materiel Center's Air Crew Equipment Laboratory. The first study [20, 26] was conducted for a period of 7 days in a small confinement chamber (450 ft^3) and involved a group of six subjects assigned to individual schedules that required 10 hr (in blocks of either 4-2-4 or 4-6 hr) of duty and permitted one 8-hr sleep period each day. In a second study [10], six subjects were confined to the same chamber and were kept on a schedule of 8 hr awake and 4 hr asleep. Of the 16 waking hours, approximately 11 were spent on duty at the work stations. A variety of tasks was presented to the subjects, but there was a difference in the combinations used for the two studies. Among the tasks included were tests of auditory-motor tracking, visual-motor tracking, time estimation, arithmetical reasoning, warning-lights monitoring, and association learning. Bioelectric measures of heart rate, respiration rate, forehead-skin temperature, and skin conductance were obtained in the second (8-day) study [25]. Results of both studies indicated that the groups of men were able to operate satisfactorily, with no marked change or deterioration in performance or physiological reactions.

A series of studies reported by Adams and Chiles [2, 3] deals with systematic variation of the work-rest schedule and the work-rest ratio for periods longer than 24 hr. The first study in this series examined 96 hr of performance on four different schedules, 2-2, 4-4, 6-6, and 8-8. As mentioned earlier, the shorter cycles, 2-2 and 4-4, were preferred by most subjects, and there was a suggestion that prolongation of the 6-6 and 8-8 cycles would lead to decrements in performance because of the boredom induced by the tasks. This study was followed by two additional 96-hr experiments, one with a 4-2 schedule and one with a 6-2 schedule. Although the performance data did not indicate that the 4-2 schedule was superior, the questionnaire data and especially the amount of sleep that subjects were able to get suggested that pronounced decrements would probably result from prolongation of the experimental period in the case of the 6-2 schedule, but not in the case of the 4-2 schedule.

In the final study of this series, operational personnel (two B-52 crews) were confined to a simulated crew compartment [1, 43] and tested on a 4-2 schedule over a 360-hr period (15 days). Under these conditions three men, with appropriate cross training, covered two positions continuously. The first crew to be tested consisted of five men, the second

consisted of six men, and both crews ranged in age from twenty-six to forty-three years. During the 4-hr work periods per day, each subject was presented with five tasks designed to measure functions such as mental computation, pattern discrimination, monitoring, and vigilance. A basic 2-hr task schedule consisting of a 30-min low-performance period (two tasks presented simultaneously) and a 90-min high-performance period (four tasks presented simultaneously) was repeated 180 times during each of the two 15-day tests. From the subject's point of view, however, there was no break in repetitions of the schedule, since two of the tasks were presented continuously at each station. Recordings of skin resistance, heart rate, respiration rate, and skin temperature were made concurrently with task performance.

Performance curves for four of the tasks are shown in Figure 3-4. These plots are based on the mean value of scores obtained from all 11 subjects and were selected as being typical of the complete data. Performance of the arithmetic-computation and meter-monitoring tasks showed a significant deterioration over the 15-day period. In contrast, performance on the pattern-discrimination task improved significantly, and the latency in responding to the occurrence of a warning-light signal was unaffected. Since these curves represent group trends, they do not reflect the fairly wide differences in performance that were observed among individual subjects. Differences in individual trends similar to those shown in Figure 3-2 were not uncommon. For example, only 2 of the subjects showed a significant decrement on all tasks, whereas 8 showed a significant improvement or no change in the performance of two or more tasks.

One of the most prominent features of the curves shown in Figure 3-4 is the reasonably consistent within-day change, or diurnal variation, in performance over the 15-day period. This is shown more clearly in Figure 3-5, where the performance of all subjects has been combined on the basis of the eight 2-hr work periods and plotted for the first, middle, and last 5-day blocks of experimentation.

The results obtained for two of the physiological measures, skin-resistance level and heart-rate level, are shown in Figure 3-6. In both cases the group trends are significant over days and in the direction of a decrease in level of autonomic activation (i.e., an increase in skin resistance and a decrease in heart rate). The prominent diurnal variation observed in both measures, plotted in 5-day blocks, is shown in Figure 3-7.

A comparison of the curves in Figures 3-5 and 3-7 indicates that the levels of performance efficiency and autonomic activation both show a well-marked diurnal rhythm. It should be pointed out, incidentally, that the ordinate value of each point on all curves is based on scores from all subjects and the abscissa value is the average of the individual

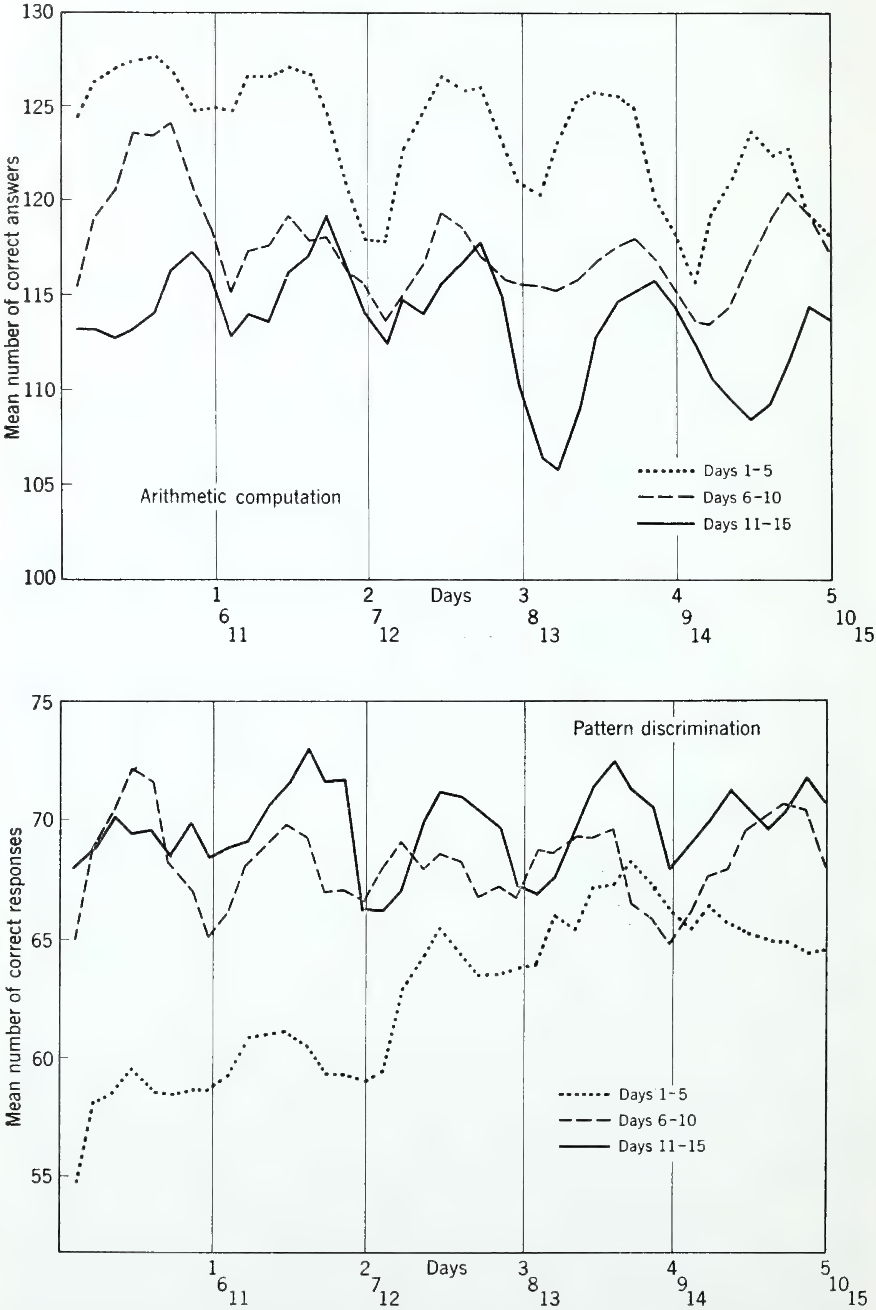


FIG. 3-4 *Level of performance achieved on four tasks by subjects committed to a continuous schedule of 4 hr work and 2 hr rest for 15 consecutive days.*

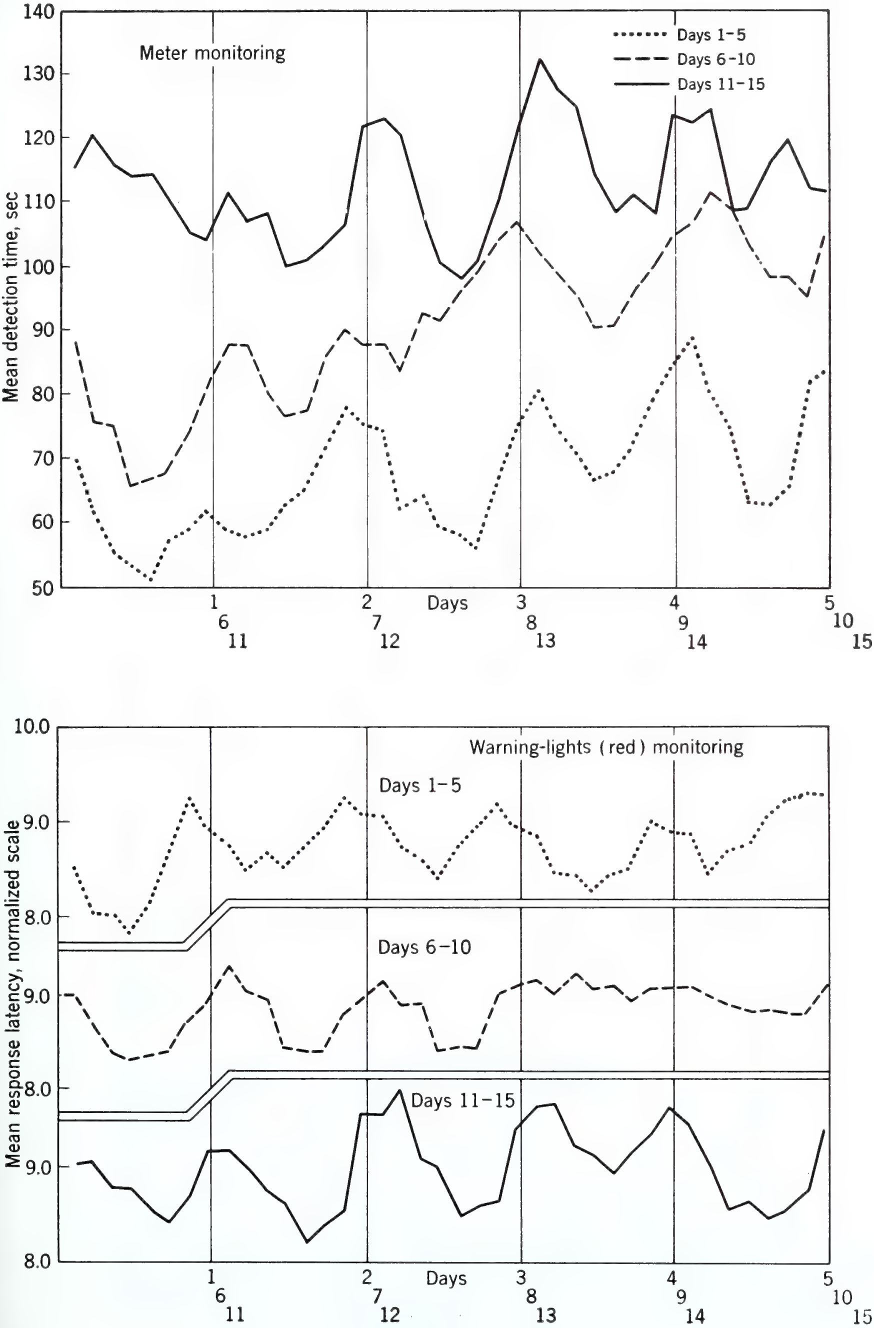


FIG. 3-4 (*continued*)

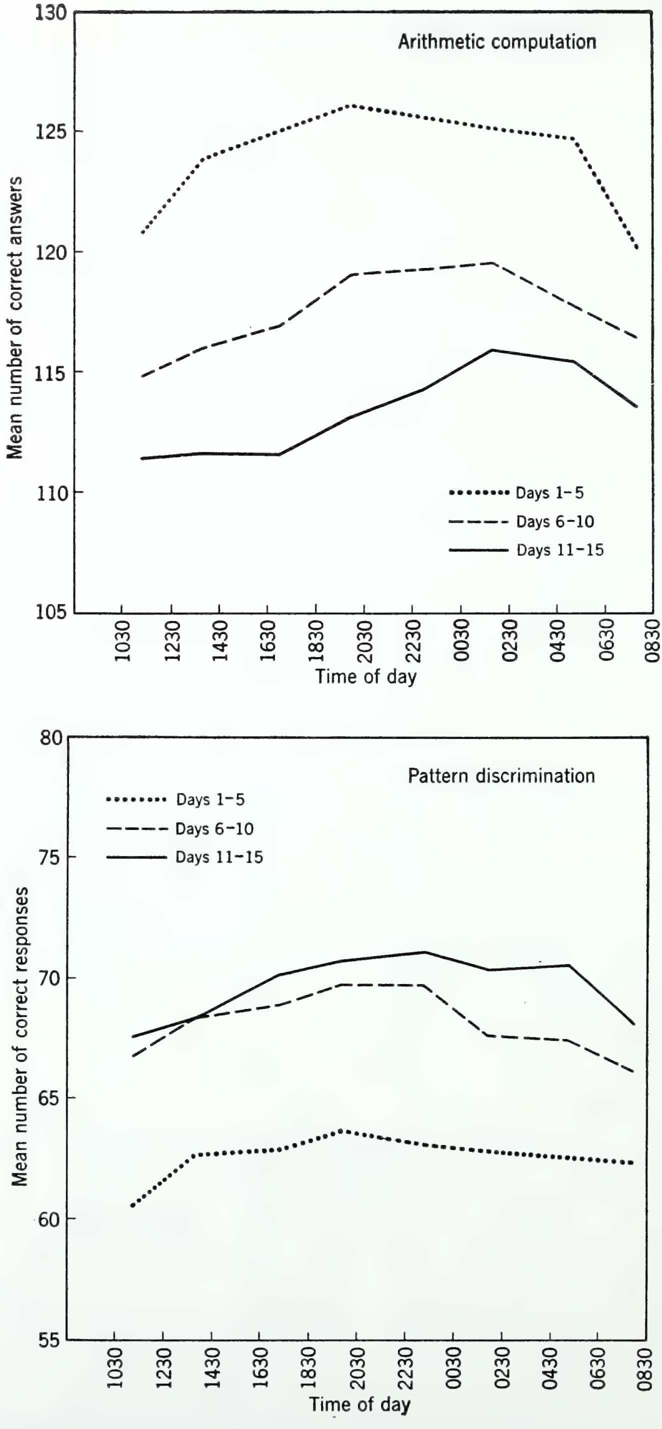
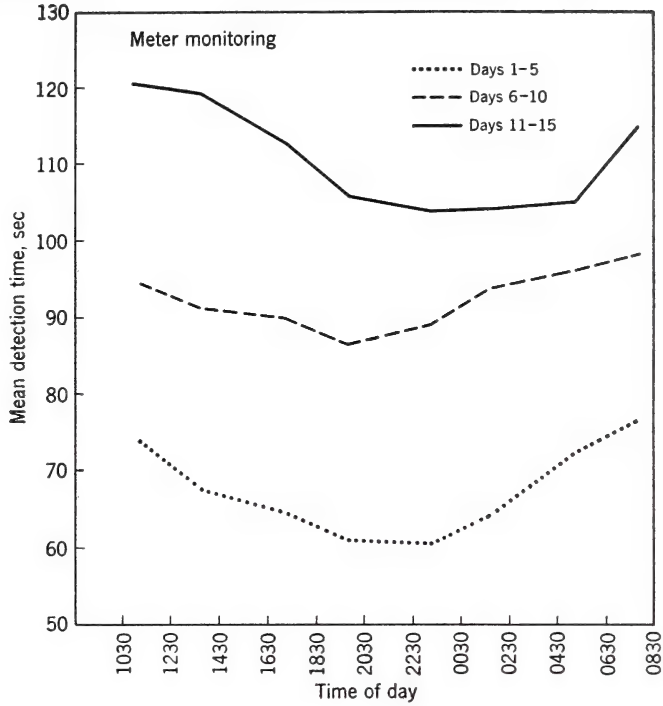


FIG. 3-5 *Within-day changes (diurnal variation) in level of task performance.*



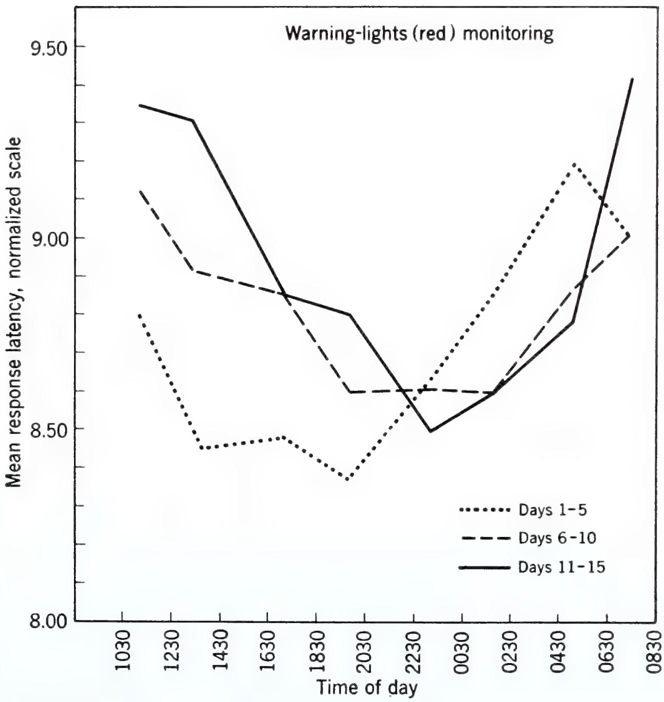


FIG. 3-5 *(continued)*

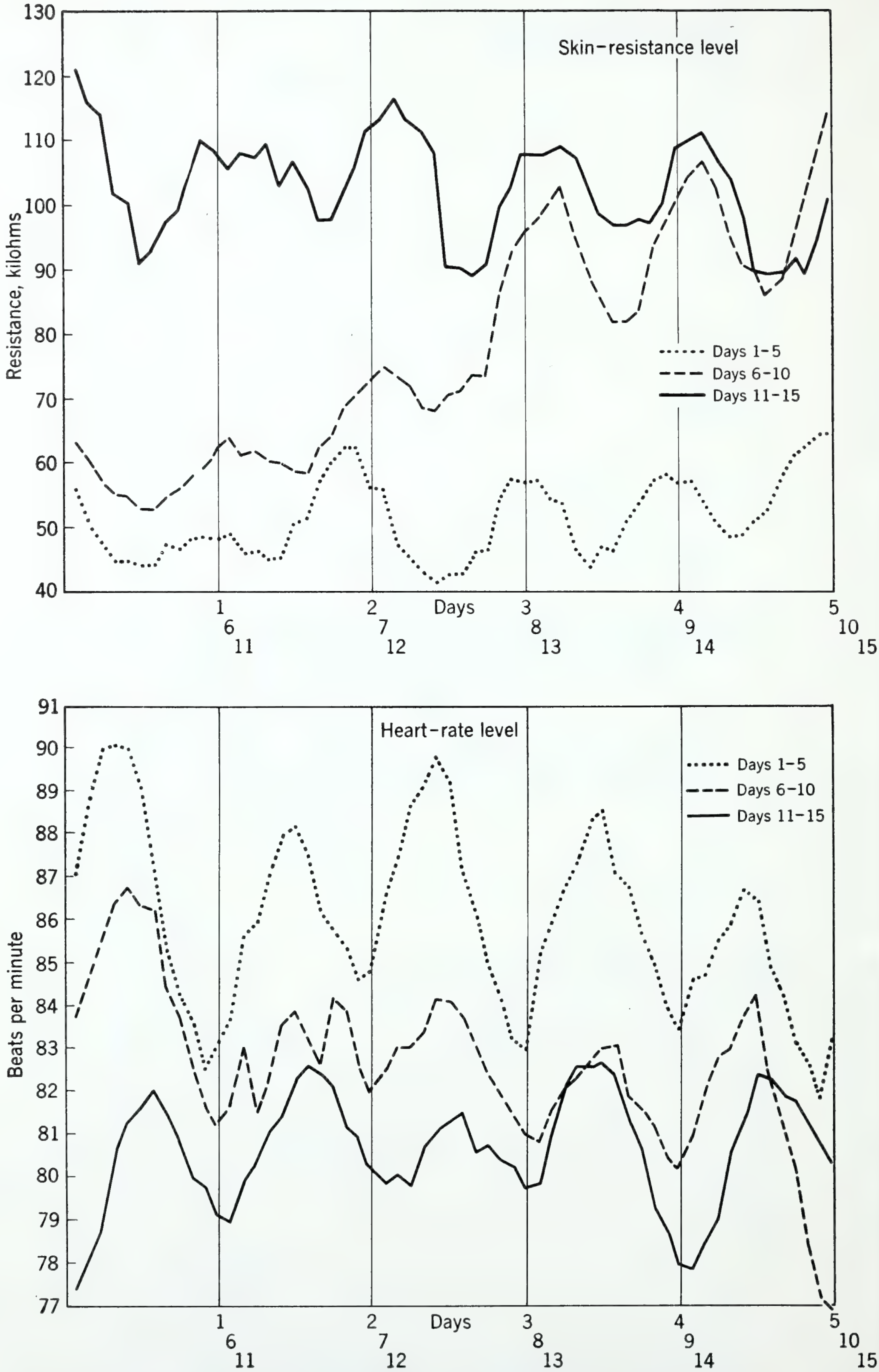
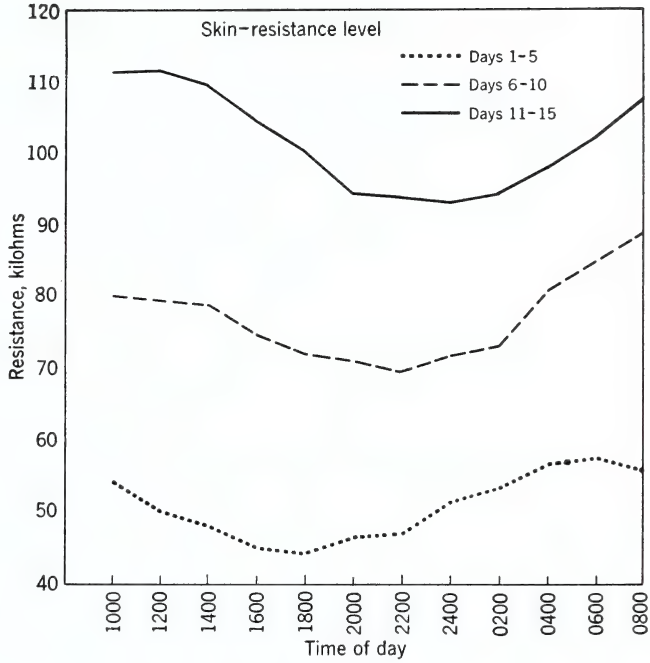


FIG. 3-6 *Skin resistance and heart-rate levels of subjects committed to a continuous schedule of 4 hr work and 2 hr rest for 15 consecutive days. Recordings made during period in which subjects performed arithmetic computation and three monitoring tasks concurrently.*



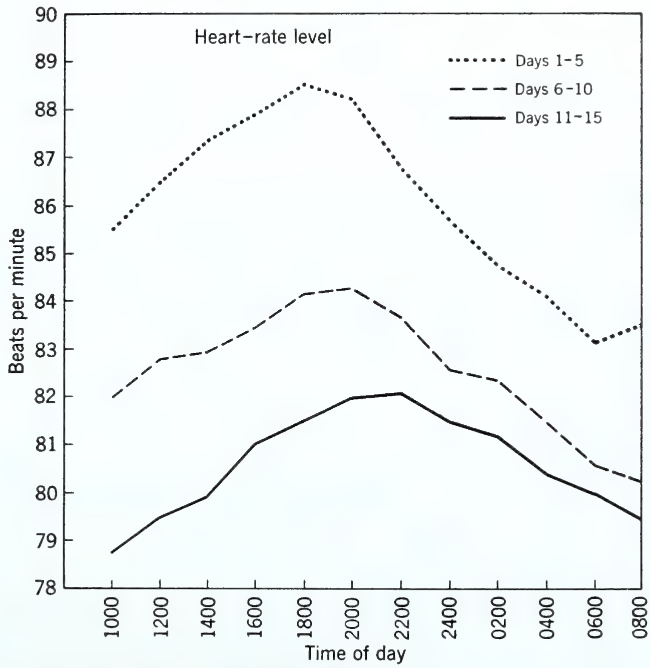


FIG. 3-7 *Within-day changes (diurnal variation) in skin resistance and heart-rate levels.*

times of measurement. Each curve, therefore, can be interpreted as representing the around-the-clock status of the system. Although the 24-hr rhythm appears to diminish somewhat during the latter portion of the 15-day period, it obviously remains as a distinct periodicity. Whether or not appropriate selection and training techniques could be used to minimize the diurnal effect under this particular work-rest schedule is, of course, not known. The question is of sufficient importance, however, to warrant further investigation.[1]

The general conclusion drawn from this series of studies is that, with a minimum amount of selection, highly motivated subjects can be found who will maintain acceptable performance levels on a schedule of 4 hr work and 2 hr rest for periods as long as 15 days and probably for 30 days. This is substantiated by several observations made during the course of the 15-day studies. First, the test conditions appeared to have less adverse effect on the performance of those subjects who were judged to have the highest levels of motivation. In this connection it should be mentioned that both crews were unselected and only one had volunteered for the test. The group-performance level for the volunteer crew was in all respects higher than the level for the nonvolunteer crew, which had been recalled from annual leave in order to participate in the study. Second, 7 of the 11 subjects showed a significant improvement in performing a learning task (pattern discrimination). The possibility that a less stringent schedule would produce more rapid improvement is not substantiated by the results of a subsequent control study. It was found that performance on the same task by 6 subjects working 4 hr/day, 5 days a week, for 6 weeks was not significantly different from that achieved by the volunteer crew working on the 4-2 schedule for an equivalent duration of task performance (120 hr). Third, in a posttest interview, the majority of the subjects indicated that they could have continued the test for at least another 15 days if it had been necessary to do so.

BIBLIOGRAPHY

1. Adams, O. S.: Aircrew Fatigue Problems during Extended Endurance Flight. Phase I: Planning, *USAF WADC Tech. Rept.* 57-510, May, 1958.

[1] The fairly consistent tendency for the times of highest within-day level of performance and autonomic activation to shift from the early evening to the late evening or early morning hours during successive days may be an artifact. Both experimental studies were begun at 0930 hr, which was approximately 3½ hr after waking. As the study progressed, 0930 was used as a reference point for the beginning of another day and psychologically may have corresponded more closely to the subject's normal hour of arising. In this sense, the subjects may have shifted to a "time zone" that differed from that of the external environment and thus made an adequate adjustment.

2. ——— and W. D. Chiles: Human Performance as a Function of the Work-Rest Cycle, *USAF WADD Tech. Rept.* 60-248, March, 1960.

3. ——— and ———: Human Performance as a Function of the Work-Rest Ratio During Prolonged Confinement, *USAF ASD Tech. Rept.* 61–720, November, 1961.

4. Arai, T.: Mental Fatigue, *Teachers Coll. Contribs. Educ.*, no. 54, 1912.

5. Ax, A. F., et al.: Quantitative Effects of Sleep Deprivation, *Am. Psychologist*, vol. 9, p. 324, 1954. (Abstract.)

6. Balke, B.: Experimental Studies on the Conditioning of Man for Space Flights, *Air Univ. Quart. Rev.*, vol. 11, no. 1, pp. 61–75, 1959.

7. Bliss, E. L., L. D. Clark, and C. D. West: Studies of Sleep Deprivation—Relationship to Schizophrenia, *Am. Med. Assoc. Arch. Neurol. Psychiat.*, vol. 81, pp. 348–359, 1959.

8. Brindley, G. S.: Intrinsic 24-Hour-Rhythms in Human Physiology and Their Relevance to the Planning of Working Programmes, *RAF Inst. Aviation Med. Flying Personnel Research Comm. Rept.* 871, 1954.

9. Bulban, E. J.: Vivid Hallucinations Plague Test Subjects, *Aviation Week*, vol. 72, pp. 57–65, Feb. 1, 1960.

10. Burns, N. M.: Environmental Requirements of Sealed Cabins for Space and Orbital Flights—A Second Study. Part 1: Rationale and Habitability Aspects of Confinement Study, *Naval Air Materiel Center Rept.* NAMC-ACEL-413, 1959.

11. Chiles, W. D.: Experimental Studies of Prolonged Wakefulness, *USAF WADC Tech. Rept.* 55-395, 1955.

12. Christensen, J. M.: Aerial Analysis of Navigator Duties with Special Reference to Equipment Design and Workplace Layout: II. Analysis of Navigator Duties under Arctic Conditions, *USAF AMC Mem. Rept.* MCREXD-694-15A, 1948.

13. Clark, R. E., et al.: The Effects of Sleep Loss on Performance of a Complex Task, *OSRD Rept.* 3153, 1943.

14. Edwards, A. S.: Effects of the Loss of One Hundred Hours of Sleep, *Am. J. Psychol.*, vol. 54, pp. 80–91, 1941.

15. Elliott, E.: Perception and Alertness, *Ergonomics*, vol. 3, pp. 357–364, 1960.

16. Frankmann, J. P., and J. A. Adams: Theories of Vigilance, *USAF CCDD Tech. Note* 60-25, 1960.

17. Fraser, D. C.: A Study of Deterioration of Performance in Aircrew. Interim Report. I. Validation of Techniques, *RAF Inst. Aviation Med. Flying Personnel Research Comm. Rept.* 804, 1952.

18. ——— and G. D. Samuel: Aircrew Fatigue in Long Range Maritime Reconnaissance: 10. Effects on Vigilance, *RAF Inst. Aviation Med. Flying Personnel Research Comm. Rept.* 907.10, 1956.

19. Freeman, G. L.: Compensatory Reinforcements of Muscular Tension Subsequent to Sleep Loss, *J. Exptl. Psychol.*, vol. 15, pp. 267–283, 1932.

20. Gaito, J., et al.: Environmental Requirements of Sealed Cabins for Space and Orbital Flight: Part 3. Performance and Habitability Aspects of Extended Confinement, *Naval Air Materiel Center Rept.* NAMC-ACEL-385, 1958.

21. Gerathewohl, S. J.: Work Proficiency in Space Cabin Simulator, *Aerospace Med.,* vol. 30, pp. 722–735, 1959.

22. Goldmark, J., and M. Hopkins: Studies in Industrial Physiology: Fatigue in Relation to Working Capacity. 1. Comparison of an Eight-Hour Plant and a Ten-Hour Plant, *USPHS Public Health Bull.* 106, 1920.

23. Gorham, W. A., D. B. Orr, and T. G. Trittipoe: Research on Behavior Impairment Due to Stress: An Experiment in Long-Term Performance, American Institute for Research, Washington, 1958.

24. Halberg, F.: The 24-hour Scale: A Time Dimension of Adaptive Functional Organization, *Perspectives Biol. Med.,* vol. 3, pp. 491–527, 1960.

25. Hanna, T. D., N. M. Burns, and E. Hendler: Environmental Requirements of Sealed Cabins for Space and Orbital Flights—A Second Study. Part 5: Some Physiological Measures on Confined Subjects Breathing Recycled Gases for Eight Days, *Naval Air Materiel Center Rept.* NAMC-ACEL-417, 1960.

26. ——— and J. Gaito: Performance and Habitability Aspects of Extended Confinement in Sealed Cabins, *Aerospace Med.,* vol. 31, pp. 399–406, 1960.

27. Hauty, G. T.: Human Performance in the Space Travel Environment, *Air Univ. Quart. Rev.,* vol. 10, no. 2, pp. 89–107, 1958.

28. ——— and R. B. Payne: Fatigue, Confinement, and Proficiency Decrement, in "Vistas in Astronautics," pp. 304–309, Pergamon Press, Inc., New York, 1958.

29. ——— and ———: Mitigation of Work Decrement, *J. Exptl. Psychol.,* vol. 49, pp. 60–67, 1955.

30. Husband, R. W.: The Comparative Value of Continuous Versus Interrupted Sleep, *J. Exptl. Psychol.,* vol. 18, pp. 792–796, 1935.

31. Huxtable, Z. L., M. H. White, and M. A. McCartor: A Re-performance and Re-interpretation of the Arai Experiment in Mental Fatigue with Three Subjects, *Psychol. Mon.,* vol. 59, no. 275, 1946.

32. Jackson, K. F.: Aircrew Fatigue in Long Range Maritime Reconnaissance: Pilot Performance, *RAF Inst. Aviation Med. Flying Personnel Research Comm. Rept.* 907.2, 1956.

33. Jerison, H. J., and R. Wallis: Experiments on Vigilance. II: One Clock and Three Clock Monitoring, *USAF WADC Tech. Rept.* 57-206, 1957.

34. Jones, B. F., et al.: Fatigue and Hours of Service of Interstate Truck Drivers, *USPHS Public Health Bull.* 265, 1941.

35. Jones, G. M.: Fatigue in Radio Operators during a Program of High Intensity, Long Duration, Flying, *Aerospace Med.,* vol. 31, pp. 478–484, 1960.

36. Katz, S. E., and C. Landis: Psychologic and Physiologic Phenomena during a Prolonged Vigil, *Am. Med. Assoc. Arch. Neurol. Psychiat.,* vol. 34, pp. 307–316, 1935.

37. Kleitman, N.: Biological Rhythms and Cycles, *Physiol. Rev.,* vol. 29, no. 1, pp. 1–30, 1949.

38. ———: "Sleep and Wakefulness," University of Chicago Press, Chicago, 1939.

39. ———: The Sleep-Wakefulness Cycle of Submarine Personnel, in Panel on Psychology and Physiology, "A Survey Report on Human Factors in

Undersea Warfare," pp. 329–341, National Research Council Committee on Undersea Warfare, Washington, 1949.

40. ——— and D. P. Jackson: Body Temperature and Performance under Different Routines, *J. Appl. Physiol.*, vol. 3, pp. 309–328, 1950.

41. ——— and E. Kleitman: Effect of Non-Twenty-Four-Hour Routines of Living on Oral Temperature and Heart Rate, *J. Appl. Physiol.*, vol 6, pp. 283–291, 1953.

42. ——— and J. Schreider: Sleepiness and Diplopia, *Am. J. Physiol.*, vol. 129, p. 398, 1940. (Abstract.)

43. Kraft, J. A.: Measurement of Stress and Fatigue in Flight Crews during Confinement, *Aerospace Med.*, vol. 30, pp. 424–430, 1959.

44. Laird, D. A., and W. Wheeler: What It Costs to Lose Sleep, *Ind. Psychol.*, vol. 1, pp. 694–696, 1926.

45. Lewis, P. R., and M. C. Lobban: Dissociation of Diurnal Rhythms in Human Subjects Living on Abnormal Time Routines, *Quart. J. Exptl. Physiol.*, vol. 42, pp. 371–386, 1957.

46. ——— and ———: Effects of Prolonged Periods of Life on Abnormal Time Routines upon Excretory Rhythms in Human Subjects, *Quart. J. Exptl. Physiol.*, vol. 42, pp. 356–371, 1957.

47. Lindsley, D. B., et al.: Radar Operator "Fatigue": The Effect of Length and Repetition of Operating Periods on Efficiency of Performance, *Natl. Defense Research Comm. Appl. Psychol. Panel Rept.* 6, 1944.

48. Luby, E. D., et al.: Sleep Deprivation: Effects on Behavior, Thinking, Motor Performance, and Biological Energy Transfer Systems, *Psychosomat. Med.*, vol. 22, pp. 182–192, 1960.

49. McIntosh, B. B., J. L. Milton, and E. L. Cole: Pilot Performance during Extended Periods of Instrument Flight, *USAF WADC Tech. Rept.* 6725, 1952.

50. Mackworth, N. H.: Researches on the Measurement of Human Performance, *Med. Research Council Spec. Rept.* 268, H.M. Stationery Office, London, 1950.

51. Mason, H. L., et al.: Result of Administration of Anterior Pituitary Adrenocorticotropic Hormone to a Normal Human Subject, *J. Clin. Endocrinol.*, vol. 8, no. 1, pp. 1–14, 1948.

52. Murray, E. J., H. L. Williams, and A. Lubin: Body Temperature and Psychological Ratings during Sleep Deprivation, *J. Exptl. Psychol.*, vol. 56, pp. 271–273, 1958.

53. Payne, R. B., and G. T. Hauty: Effect of Psychological Feedback upon Work Decrement, *J. Exptl. Psychol.*, vol. 50, pp. 343–351, 1955.

54. ——— and ———: The Effects of Experimentally Induced Attitudes upon Task Proficiency, *J. Exptl. Psychol.*, vol. 47, pp. 267–273, 1954.

55. Ray, J. T., O. E. Martin, and E. A. Alluisi: Human Performance as a Function of the Work-Rest Cycle: A Review of Selected Studies, *Natl. Acad. Sci.–Natl. Research Council Publ.* 882, 1961.

56. Steinkamp, G. R., et al.: Human Experimentation in the Space Cabin Simulator: Development of Life Support Systems and Results of Initial Seven-Day Flight, *USAF School Aviation Med. Rept.* 59-101, 1959.

57. Strughold, H.: The Physiological Day-Night Cycle in Global Flights, *J. Aviation Med.*, vol. 23, pp. 464–473, 1952.

58. Tyler, D. B.: Effect of Amphetamine Sulfate and Some Barbiturates on the Fatigue Produced by Prolonged Wakefulness, *Am. J. Physiol.*, vol. 150, pp. 253–262, 1947.

59. Webb, W. B., and R. J. Wherry: Vigilance in Prolonged and Repeated Sessions, *Perceptual Motor Skills*, vol. 10, pp. 111–114, 1960.

60. Welford, A. T., R. A. Brown, and J. E. Gabb: Two Experiments on Fatigue as Affecting Skilled Performance in Civilian Air Crew, *Brit. J. Psychol.*, vol. 40, pp. 195–211, 1950.

Chapter 4

ACCELERATION AND BODY DISTORTION

*Carl Clark**

TOLERANCE TO ACCELERATIONS OTHER THAN IMPACT FOLLOWING A FREE fall is greatly dependent on the form of restraint system used. Acceleration-time tolerance curves are drawn with careful distinctions made among the different orientations of the acceleration vector with respect to the seat but with inadequate distinctions made among the different forms of seats and other aspects of the restraint. Tolerance to acceleration depends more on the extent of body distortion, a consequence of the interaction of the reactive forces caused by acceleration and the opposing forces caused by tissue displacements and the restraint system, than on the reactive force due to acceleration alone. Damage is done, not by the force resulting from acceleration, but by the distortions produced as a consequence of this force.

What acceleration can a human being withstand? The "ultimate" values are unknown, but surely they are higher than what are now experimented with. DeHaven reports human survivals of accidental falls with impacts estimated at over 150 G [24]. Margaria reports survival of mouse embryos (with no air spaces) at impacts of 10,000 G [44].

ACCELERATION TERMINOLOGY

Recently efforts have been made to propose a consistent physiological-acceleration terminology [17]. These are illustrated in Figure 4-1 and summarized as follows:

1. The unit for the physiological acceleration shall be G so that this acceleration will be distinguished from the "true" displacement acceleration, generally designated by aerodynamicists by the unit g.[1]

* Life Sciences Department, The Martin Company, Baltimore, Md.

A major part of the work described in this chapter was done while the author was with the Biophysics and Bioastronautics Division of the Aviation Medical Acceleration Laboratory, U.S. Naval Air Development Center, Johnsville, Pa.

[1] The acceleration due to gravity, approximately 9.80 m/sec^2 (32.2 ft/sec^2 or 22.0 mph/sec).

The physiological acceleration is the total reactive force divided by the body mass; hence it includes effects of both displacement and resisted gravitational acceleration.

2. The physiological-acceleration axes represent directions of the reactive displacements of organs and tissues with respect to the skeleton. The Z axis is down the spine, with $+G_z$ (unit vector) designations for accelerations causing the heart and other body parts to displace downward (caudally). The X axis is front to back, with $+G_x$ designations for accelerations causing the heart to be displaced back toward the spine (dorsally). The Y axis is right to left, with $+G_y$ designations for accelerations causing the heart to be displaced to the left. For accelerations in which effects on the entire body are of concern, the origin of the axes shall be halfway between the anterior (rostral) surfaces of the iliac crests, with the Z axis passing through the midpoint between the suprasternal notch and the dorsal surface of the dorsal spine of the last cervical vertebra to this origin. The X and Y axes are mutually perpendic-

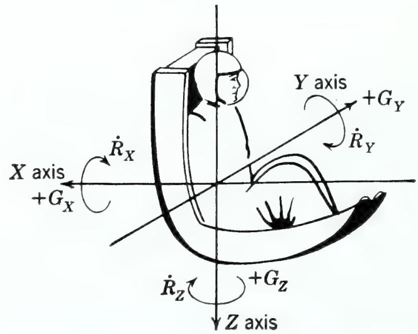


FIG. 4-1 *Physiological description of acceleration. (Official U.S. Navy photograph.)*

Description of Heart Motion
(Directions are those of heart displacement with respect to the skeleton.)

LINEAR ACCELERATION MODES

Actual	Other description		Unit vector
Toward spine	Eyeballs-in	Chest-to-back	$+G_x$
Toward sternum	Eyeballs-out	Back-to-chest	$-G_x$
Toward feet	Eyeballs-down	Head-to-foot	$+G_z$
Toward head	Eyeballs-up	Foot-to-head	$-G_z$
Toward left	Eyeballs-left		$+G_y$
Toward right	Eyeballs-right		$-G_y$

ANGULAR ACCELERATION MODES

Acceleration about X axis (The heart rolls left in the chest.)	$+\dot{R}_x$
Acceleration about Y axis (The heart pitches down.)	$+\dot{R}_y$
Acceleration about Z axis (The heart yaws left.)	$+\dot{R}_z$

ular to this Z axis. For acceleration effects on the vestibular apparatus, the origin of the head axes shall be the midpoint between the external auditory meatuses (on the Y axis), with the X axis passing from the ventral medial margin of the nasal bones through this origin.

3. Angular accelerations that cause the heart to rotate (roll) to the left within the skeleton shall be specified by the $\ddot{R}_x$ unit vector, representing radians per second per second about the X axis. Angular velocities in the same sense shall be specified by the $+\dot{R}_x$ unit vector, representing radians per second about the X axis. Similarly, $+\dot{R}_y$ represents pitch down of the heart within the skeleton, and $+\dot{R}_z$ represents yaw right of the heart within the skeleton.

4. Linear-acceleration environments may be represented by the three acceleration components (along the G_x, G_y, and G_z unit vectors) or by a resultant acceleration and the azimuth and altitude angles of the resultant with respect to the body axes. Azimuth is measured from the $+X$ axis (to the back), with positive rotation clockwise as seen from above. Altitude is measured from the horizontal (XY) plane, with positive angles when in the hemisphere of the $+Z$ axis (downward). Thus a man reclining in a chair tipped back 45° experiences 0.7 G_x and 0.7 G_z, or 1 G/0°, 45°.

5. Whenever rotations accompany linear accelerations, the reference point for the linear accelerations should be specified and the time histories of the angular velocities and angular accelerations should accompany the time histories of the linear accelerations in order that linear accelerations at other points may be computed.

THE ACCELERATION ENVIRONMENT

Table 4-1 lists the g seconds[2] of various travel speeds. In this, the early "slingshot" era of rocket travel, in which the peak velocity is attained within a few minutes and is followed by a coasting period, the g seconds utilized by man for actual travel have thus far barely exceeded 1,000 g sec. However, with the development of advanced power sources, including nuclear power, longer-duration accelerations and higher speeds will be attained, although even the centrifuge simulation of speeds of several g months will probably not be carried out for several years.

Although it is reasonable to expect with the development of improved flight sources of energy that design acceleration magnitudes may decrease

[2] A g second is displacement acceleration determined by g times the time of acceleration in seconds.

for future vehicles and give lower accelerations for longer durations, it is also clear that, with the higher speeds attained, higher accelerations may occur under emergency situations. The extent to which rocket vehicles should be designed for such emergencies is now a matter of controversy, analogous to the controversy over design-limit loads for airplanes [3]. To what extent should the design use of the vehicle be jeopardized by the addition of weight to the vehicle for survival in emergency use? With the older restraint systems, which consisted of

TABLE 4-1 *Various Travel Speeds (Seconds)*

Condition	Speed		g sec
	Metric	English	
Fall 4.9 m (16.1 ft)	9.8 m/sec	32.2 ft/sec (22 mph)	1
Drive an automobile	100 km/hr	62.1 mph	2.8
Fly a commercial jet airplane	1,000 km/hr	621 mph	28
Fly the X-15 rocket airplane*	1.61 km/sec	1.00 mps (3,609 mph)	163
Fly the Redstone rocket†	2.32 km/sec	1.44 mps (5,180 mph)	235
Orbit the earth (at 200 km)—tangential velocity‡	7.78 km/sec	4.84 mps	790
Escape the earth—radial velocity	11.2 km/sec	6.95 mps	1,140
Orbit the sun (at 150 × 10⁶ km)—tangential velocity	29.8 km/sec	18.5 mps	3,020
Escape the sun (from 150 × 10⁶ km)—radial velocity	42.4 km/sec	26.4 mps	4,320

* Robert White, X-15 flight, June 24, 1961.
† Alan Shepard, Mercury Freedom 7 flight, May 5, 1961.
‡ Yuri Gagarin, Vostok flight, Apr. 12, 1961.
NOTE: To attain altitude near the earth requires 143 g sec/100 km or 180 g sec/100 miles. To attain 10^6 km/hr requires 7.85 g hr. To attain 10^6 mph requires 12.6 g hr. To attain one-half the speed of light requires 5.9 g months.

lap belts and shoulder straps, vehicle failure might be acceptable, since the crew usually lost control capabilities for accelerations lasting tens of seconds outside the range of perhaps +9 to −5 G_z, +15 to −15 G_x, and ±6 G_y. However, with the development of improved restraint and protective systems, this letting the vehicle fail simply because the crew has already lost control no longer applies. In the introduction of commercial rocket travel the trend might be toward greater acceleration capability for the vehicle and an improved restraint-protective system for emergency use, as well as toward reduction of the probability of

the occurrence of high-acceleration events. It is, therefore, not too early to begin research on rocket-vehicle emergency and crash injury.

GROSS BODY DISTORTIONS

The body musculature can provide some support against the inertial forces accompanying acceleration. Figure 4-2 specifies the vehicle accelerations at which it is just possible to make head or limb motions in various directions or to prevent involuntary motions when the body is tensed [16]. These values are preliminary and are, of course, affected by the strength of the subject and the weight of equipment carried. However, they do determine the minimum restraint required for preventing involuntary head or limb motion. It has been found that in practice it is desirable to provide restraint well before these limits are reached, so that the pilots can relax in the restraints and carry out their control tasks without the great muscular effort necessary to maintain head and limb positions. Under conditions of oscillating accelerations, it is even more difficult to prevent involuntary motions, which may result in involuntary pilot-control inputs, which may excite further oscillations.

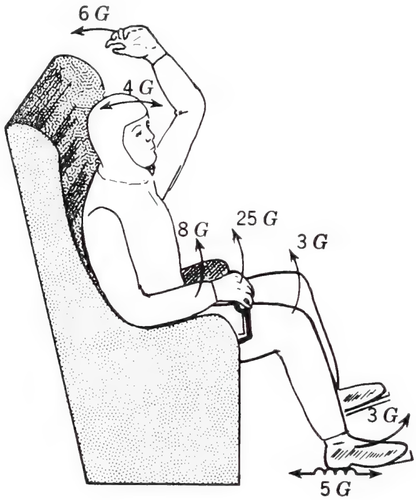


FIG. 4-2 *Body movements (indicated by arrows) that are just possible at the vehicle accelerations listed. (Official U.S. Navy photograph.)*

Simple ankle straps may be adequate for leg restraint. If the torso is restrained only with a lap belt and shoulder straps, torso support can be increased by the addition of broad thigh straps and/or straps over the knees. Stick grips at the end of the armrests can provide adequate arm restraint to beyond $+15\ G_x$ or $-11\ G_x$ or $+11\ G_z$ if a control task is not being done. Restraints of the upper arm and wrist are desirable if fine control motions are to be made by using a "console stick." In jostle environments, coupling of the hand to the control might be improved and involuntary control inputs reduced by immersing the hand and console control in water. However, no study of this approach has been performed.

Head restraint is desirable for accelerations above ± 2 G_x when the acceleration lasts longer than it takes the head to swing to the limits of its travel. For most people, head restraint is essential above about ± 4 G_x. Likewise, unless great care is taken not to flex the neck, loads above about $+5$ G_z require head restraint, particularly when combined with loads above -2 G_x that initiate flexion. With lateral accelerations of slow onset (about 0.05 G/sec), the head comfortably rotates (yaws and pitches down) so that even 5 G_y does not produce painful neck motion [13]. For $-G_x$ accelerations, above perhaps -5 G_x, restraint of the helmet may not provide adequate head restraint, for the head slides forward against the top front of the helmet and onto the chin strap. Development of an individually molded "cheekpiece" restraint in addition to the chin strap and of a special helmet with broad facial contact has made possible sustaining -10 G_x for 90 sec without particular discomfort [57].

The form of head restraint is a matter of continuing concern in the effort to avoid differing displacements of the head and body. The solution for the X-15 rocket aircraft is that of having the helmet in a "socket" and using lateral top and retractable front head "bumpers" that allow room for head motion if the torso moves within its restraints [16]. In Project Mercury, the helmet is within a molded or contoured socket and allows good lateral (G_y) and $+G_x$ support [59]. No special support is provided for the emergency condition of $-G_x$ of the low-altitude abort cases; the helmet and neck ring support the head and limit neck deflection. In 1957 it was suggested that the head could be restrained by supporting it on the shoulders and chest by using a "chin bumper" and helmet tightened onto the shoulders to prevent head motion [16]. Then restraint of the torso would also support the head, with no chance of differential motion.

Distortions of the eye under acceleration may contribute to reduced visual capability, which also may be partly caused by reduced blood supply. Motions of the eyeball in the head, with a resonance frequency of between 60 and 90 cps [32], and distortions of the eyeball shape might be reduced by water goggles [34]. Resonance of the head, such that head amplitudes may exceed shoulder amplitudes by a factor of 3, occurs in the range of 20 to 30 cps [32].

Torso-restraint systems vary from the seat, lap belt, and shoulder straps used by Adolphe Pégoud in the first intentional inverted-airplane flight in 1913 [17, 56] to the deeply indented "contour couch" suggested for development by Maxime Faget, National Advisory Committee for Aeronautics, and used by Carter Collins in December, 1958, to reach $+25$ G_x as the peak of a versine waveform with a period of 40 sec [16, 21]. The conventional lap belt, shoulder straps, and airplane seat

provide incomplete restraint for lateral accelerations, allowing displacements of 5 to 7 in. at 4 G_y [13]. Improved lateral restraint may be provided by armrests or preferably by contouring the seat and back about the body. For $-G_x$ accelerations ("eyeballs out") broad straps over the chest and pelvis [16, 52] provide a more comfortable support and less body distortion than is provided by either shoulder straps and lap belt or both plus a chest strap. It is desirable not to restrain the entire abdomen in addition to the chest, for if at least part of the abdomen is left unrestrained, diaphragmatic breathing can, with training, be maintained even at high acceleration.

A restraint with "quick release" advantages beyond those of the back-contour couch is a backpiece, worn by the pilot, with projecting knobs that are secured in holes in the cockpit "hardback" [57]. A rubber bladder inside the backpiece is inflated with air or water to a comfortable level and is used to push the subject against a front vest. There is also a rubber bladder between the pilot's seat and legs and inside the secured leg pieces. A high-pressure inflated restraint has been developed [1] that avoids the difficulties of bladder deformation and low-frequency resonance by using bladders traversed by internal strings. Pressures to 50 psi may be used in the bladders without application of pressure to the subject.

A "box restraint" in which the front straps of a back-contoured couch are replaced by a front-contoured support, thus allowing the back-contoured couch to close, is under development by the author. An alternative form of broad-contact restraint proposed in 1960, in which the subject is "foamed" in place by a liquid that reacts with a catalyst to form a rigid polyurethane foam of 2 to 6 lb/ft^3 density, is being developed under the direction of R. Flanagan Gray [58] and has been tested by the author. Subsequently another subject was similarly foamed and rode the centrifuge at 6 G_y. He reported that the foam provided an excellent support, without the painful clavicle pressure point that develops at 4 G_y when lap belt and shoulder straps are worn.

FLUID AND TISSUE DISPLACEMENTS

The previously discussed restraint procedures can adequately hold the body components in place during acceleration. Nevertheless, if the body components are not held to constant volume, blood can be displaced into or from them during this period and internal distortions and redistributions can occur.

In 1897 Hill and Bernard showed the significant displacement of blood into the splanchnic circulation by bringing an animal upright at

1 G [39]. They showed the effect of an abdominal belt or support in reducing this blood displacement and accompanying circulatory adjustments. Thurston [17, 55], in 1903, may have been the first human being to become unconscious under centrifugation, at 6.47 G_z in a Maxim "roundabout," or captive flying machine. The early flight occurrences [17] of loss of vision and consciousness during acrobatics were at first attributed to vestibular action [36] and later to circulatory effects [4, 23, 26]. In the same period when Poppen was developing his abdominal belt [50, 51], Marshall [45] recommended a pilot belt bracing the entire abdomen for preventing "anemia of the eyes"; the belt was to be filled by an air scoop opened when acceleration deflected a weighted valve. Further developments of the antiblackout suit, or G suit, followed, with the demonstration that "positive" acceleration ($+G_z$) produces a reduction in blood pressure at eye level, which leads to a reduction in peripheral vision (grayout), a complete loss of vision (blackout), and finally unconsciousness [62, 66]. A pressure difference of 100 mm Hg along a blood column is equivalent to 130 G cm of blood column along the resultant acceleration vector. Since physiological reflexes tend to maintain an approximately uniform central pressure gradient across certain central blood-vessel walls, organs along the $+G_z$ vector would have a reduced blood pressure, which can be estimated when the G and centimeters along the vector are known. For a person with a blackout level of about 4 G_z when relaxed and "unprotected," use of the antiblackout suit may give a relaxed blackout level of about 5.5 G_z on a centrifuge or perhaps 7 G_z with the excitement of actual flight.

Muscular effort may also be used to reduce blood pooling. Orlebar noted that in doing turns some pilots found benefit in tightening up belly muscles so long as they avoided holding their breaths [48]. Experimental studies indicated that abdominal straining with a closed glottis, achieved by elevating thoracic pressure, initially led to an elevated peripheral pressure for a heartbeat or two but thereafter reduced venous return when under the moderate accelerations of aircraft flight [61, 65]. This led to a falling central blood pressure. Straining with an open glottis or with the glottis closed only for short periods provides about 1 G_z protection. MacWilliam recommended binding the legs or keeping them in motion for reducing blood pooling there [42]. Tensing of the muscles can also serve this function.

As the powered human-centrifuge work began, it was rapidly confirmed that acceleration tolerance was increased by placing the body in a position transverse to the acceleration vector, that is, by shortening the blood column acted upon [2, 8, 25]. Tolerance then became a matter of breathing difficulty and chest discomfort. It is interesting that both Bührlen [8] and Armstrong and Heim [2] note that it is easiest to breath

when the acceleration vector passes side to side through the body (G_y), although Bührlen next favors the prone position ($-G_x$), whereas Armstrong and Heim next favor the supine position ($+G_x$). Recently the author, John Weaver, and R. F. Gray reexamined lateral-G tolerance with a better restraint than had been used in an earlier study [13]. In one series of runs to 9 G_y (versine waveform of 30 sec period) Gray felt that respiration was easier than it had been in the experiment with $+9$ G_x and -9 G_x. The author reported grayout in the "inboard" eye at 6 G_y and above. At 8 G_y he experienced a peculiar reticulation of the visual field, a phenomenon unfamiliar to Dr. Thomas Duane, an ophthalmologist with considerable centrifuge experience. Foveal vision was clear, as were points in a rough grid approximately 1 cm apart at a 30-cm viewing distance, but intervening areas were obscured by white "whirlpools." This condition lasted for about 15 min after the run, clearing first with the inboard eye. Retinal edema was not apparent. Thereafter, clear vision returned. Two days later, after detailed visual-study preparations, the author rode to 9 G_y and experienced some chest pain, not later reported by Gray, but no visual defect. Weaver rode to 8 G_y and experienced moderate chest pain; Weaver and Gray observed no visual defects, although a subject in an earlier study [13] had experienced blurring of the "outboard" eye for about 15 min after a series of runs reaching 4 G_y. Weaver and Clark prefer $+9$ G_x to 9 G_y; it is possible that body type makes a difference here, for Gray has a larger chest and is typically an abdominal (diaphragmatic) breather. Methods of increasing a pilot's G tolerance by having him crouch [60], lie in a prone position [63], or operate a supinating seat [31, 60] have been explored.

Another early method of G protection, attributed to the Japanese and later studied in the United States, was that of taping the body with elastic bandages [41]. Taping from the feet to the xiphoid process gave 0.5 to 1 G protection, but with a discomfort not experienced in the G suits that are inflated only during acceleration. However, taping has found uses in more recent centrifuge studies [16]. In some cases, elastic stockings and tight gloves can reduce limb engorgement, petechiae, and pain.

These methods of increasing acceleration tolerance by G suits, straining, assuming new body positions, or taping may reduce blood pooling and thereby maintain venous return or reduce the blood-column height acted on by the acceleration, but blood displacement can still occur. As the dependent skin capillaries become engorged, they hemorrhage and form petechiae of pencil-dot size. According to approximate estimates, this occurs when the capillary pressure exceeds the usual pressure by 100 to 150 mm Hg (130 to 200 G cm of blood) above the local tissue pressure for a few minutes or by some higher value for a shorter time [18]. Capillary fragility is evidently a function of past ex-

perience; it is evident that repeated exposures produce decreasing numbers of petechiae, although this phenomenon has not been systematically studied on a centrifuge. Autopsies following accidents involving high acceleration show these petechiae to be distributed within the body as well as in the skin, particularly near tissues or interfaces of differing densities. High-acceleration shock waves presumably cause the local tissue distortion (by reaching the elastic limits of the capillaries) before the surrounding local tissues can provide adequate counterpressure. For slowly applied acceleration loads, however, such internal petechiae would be expected only where the counterpressure could not build up, particularly around gas-filled spaces. Postacceleration examination should include the examination for petechiae, not only of the skin, but also of the sinuses, tympanic membranes, etc. [29, 38].

Following acceleration of about 10 G (0°, 45°) or 10 G_x or -10 G_x or above, a number of people have experienced postrun vestibular disorders, such as a tendency to fall upon first standing up, that last even for several weeks. Recovery thereafter seems complete. After such accelerations, some postrun fatigue is noted, as well as an occasional moderate headache, "cured" by a night's sleep. In one case involving -5 G_y and resulting in subconjunctival hemorrhaging, a sharp pain of brief duration was elicited through increasing cranial pressure by about 40 mm Hg, by having the subject kneel and touch the floor with the top of his head [13]. The pain continued with decreasing intensity for 68 hr. Although an earlier review indicated no detectable permanent damage due to centrifuge experiments [67], means of observation for such damage were and still are not refined. Implications are that the higher and longer-duration accelerations currently studied are approaching values that might cause irreversible injury with the present means of protection. Galambos and a subcommittee of the Acceleration Panel of the Committee on Bioastronautics have studied this problem and made recommendations [29, 38].

An approximate estimate indicates that subconjunctival ("scleral") hemorrhages occur when a blood "column" of more than 200 G cm extends above the eye along the resultant G vector for several seconds or longer [13]. One case considered occurred at -5 G_y. These eye hemorrhages are not painful, and recovery seems complete in 2 weeks. Squinting of the eyelid provides some protection; more would be provided by Gray's water goggles [34]. The repeated exposures of stunt pilots to -5 G_z do not produce hemorrhages [5].

Both hyperemia and ischemia of body parts down or up along the G vector can cause some discomfort in about 20 sec, at G levels that produce petechiae in the unprotected surrounding skin or prevent blood flow. In general, the higher the acceleration, the greater the care that should be

taken to hold the limbs with minimum extension along the resultant G vector. In prerocket acceleration studies, this was less of a problem, since accelerations rarely lasted more than 20 sec.

There is still inadequate understanding of the mechanisms limiting tolerance to transverse acceleration. Earlier work [7] indicated that a back angle of 20° was desirable for reducing chest pain, which might have been due to stretching of the mediastinum. In subsequent work, levels up to $+25\ G_x$ were attained by lowering the back angle to 8° and thereby reducing the blackout that occurred at higher back angles [16, 21]. The subjects used in these experiments developed a straining technique that reduced the chest pain experienced by the earlier workers, a technique subsequently developed by the National Aeronautics and Space Administration Project Mercury Astronauts [10, 11] and others [52]. With additional straining experience, a lower back angle (0 to 5°) may increase tolerance further. It must be emphasized how few people have made these high-G centrifuge runs: Gordon Cooper reached $+18\ G_x$ and blacked out, R. Flanagan Gray has reached 21 G_x without blackout, and Carter Collins has reached 25 G_x with a considerable amount of blurring of the visual field, called "whiteout." In addition, Martin Webb reached $-28\ G_x$ under water, Carter Collins reached $-26\ G_x$ under water, R. Flanagan Gray reached $-31\ G_x$ under water, and R. Flanagan Gray reached $+16\ G_z$ in water to chest level. There is probably no higher centrifuge-acceleration experience, although higher values have been attained under impact conditions of short duration; for example, Eli Beeding has reached $+83\ G_x$ [6], John Swearingen has reached $+95\ G_z$ measured on a rigid seat (with shoulder acceleration reaching 10 G_z) [54], and John Stapp has reached over $-32\ G_x$ [53].

Since there has been so little centrifuge experience above 18 G_x, the physiological significance of the back contouring of the contour couch has not been established. A deeply indented contour couch molded to Collins [16] may have reduced chest distortion and thereby contributed to his ability to reach 25 G_x. However, this support did not prevent extensive petechiae on Collins's back. Later experiments have used less deeply indented contour couches; for example, the Mercury couches have the arms beside the chest rather than an intervening wall of the couch [59]. The present individually molded contour couches may increase comfort by minimizing local pressure points, but they do not increase ultimate tolerance to accelerations lasting less than 10 min. Indeed Gray reached 21 G_x on a hard-rubber "flat-back" couch without any contouring. The internal chest support provided by straining appears to be more important than the external chest support.

The physiological significance of leg position has also been inadequately studied. When ejection from the reentry capsule was being con-

sidered for the Mercury capsule, the legs of the contour couch were extended. Subsequently, the legs were flexed in order to reduce body length [16]. Under moderate ($+G_x$) acceleration, blood from the legs might compensate for venous pooling in the abdomen and thus increase tolerance. However, for higher accelerations or accelerations of longer duration, ischemic pain of the legs and loss of leg functions might limit tolerance. Concern may also be expressed for having the legs flexed 90° at the waist during $-G_x$ acceleration, although it is easier to limit hyperemia by elastic stockings or other devices than to compensate for the ischemia of the $+G_x$ legs-flexed situation. Until water protection can be provided, placing the limbs in a position nearly parallel to the body axis should be considered for future flight vehicles that can experience over 10 G.

In 1957, in the first Johnsville centrifuge program simulating the accelerations of a proposed rocket vehicle, a brief study was made of the benefit of inflating the antiblackout suit's abdominal and leg bladders during the $-G_x$ of the reentry by having a G-valve sense G_x in addition to the valve sensing G_z [18]. A. Scott Crossfield concluded that this gave more discomfort than aid and that straining, particularly tensing the legs, was adequate for the $-G_x$ condition, which may reach -4 G_x. The X-15 is flown with the G suit inflated in proportion only to G_z. Experiencing $+G_x$, Collins and Gray felt that, although inflation of the G suit may have reduced the need for straining below perhaps 15 G_x, at higher acceleration the G suit seemed to interfere with straining and thus reduced tolerance [16]. With proper training, straining can be more effective than the airplane type of G suit in increasing acceleration tolerance; the Mercury vehicle is flown without a G suit. A suit that can effectively replace the straining actions necessary during high transverse accelerations has yet to be built.

The straining procedure used for accelerations above $+8$ G_x with a steadily increasing acceleration pattern is as follows: First the legs are tensed. This is more easily done by pushing on rudder pedals or against leg restraints than by attempting to tense the legs by balancing antagonistic muscle actions. The arms may be similarly tensed by increasing the grip, applying pressure outward at the elbows, etc., although care must be used if a hand-control task is in progress. Next the abdomen is tensed. To involve the deep muscles, the anal sphincter is constricted and the gluteal muscles are tightened. The chest is inflated, and respiration at first continues with chest volumes above the midvalue. At higher transverse accelerations, blood begins to pool in dependent portions of the lung [37]. At lower acceleration, increasing chest pressure, by straining with a closed glottis, would lead to a decreasing venous return and a falling blood pressure. At higher acceleration, chest straining with a closed glottis for increasing the chest's internal pressure reduces the pool-

ing of blood in the lung yet maintains venous return, and hence the blood pressure is maintained or elevated. Sternal pain is found to be reduced if the chest is inflated rather than deflated during this chest-straining period. Short breaths may be attempted during this period; however, if the chest straining is let up for more than a heart beat or two at above approximately 12 G_x, chest pain or blackout may develop. Unfortunately, no adequately instrumented physiological study has been done of these straining procedures and their consequences, with pressure transducers in the rectum, stomach, esophagus, and various parts of the circulatory tree and with the detection of muscle tensions by electromyography or the measurement of local loads. Nevertheless, it should be possible to develop procedures for partially training subjects in the proper straining technique without actual centrifugation.

As the duration of acceleration above about 10 G_x increases toward and beyond the breath-holding time, more care must be used in attempting to breathe than in straining. At the same time, the efficiency of breathing may be greatly reduced by the increased weight of the chest, the altered distribution of blood, the increasing atelectasis, etc. Cherniack et al. have written an excellent review of these respiration problems [9]. Fortunately a reentry from earth orbit involves only about 1,000 G sec of acceleration time (see Table 4-1); it is unusual to have reentry conditions involving more than 60 sec at or above 8 G. At subatmospheric pressure and on 100 per cent oxygen, straining may be less effective and atelectasis may increase [14], although experience has indicated that selected and trained subjects, the Mercury astronauts, can make successful simulated reentries while breathing 100 per cent oxygen at 5 psi pressure. With higher accelerations of longer duration and with velocities of earth escape and greater, further means of protection may be necessary. Alternatively, if the motors are developed for long-duration thrust at low G, these problems may be reduced. One subject has tolerated 2 G for 24 hr (172,800 G sec) on the Johnsville centrifuge without any straining or respiratory difficulties [12, 16].

WATER PROTECTION

Lansberg [40] notes Tsiolkovski's early suggestion of water immersion as a means of acceleration protection and Oberth's [47] rejection of it. The pressure of water surrounding the body under acceleration just matches the gradient of pressure of the blood and tissues within the body; thus deformations can be prevented. Dependent portions of the body do not become engorged with blood. However, the vertebrate body is not of uniform density; localized deformations may still occur particularly

near bones and the gas-filled spaces (lungs, sinuses, intestines, etc.). An application of the water-protection technique was the Franks "flight suit," in which interconnecting water-filled bladders within a suit of low distensibility made contact with the body from heart level to the feet [22, 28]. This was of lower weight than the 1934 German "water suit" [30], in which the body was immersed from neck to feet, but it was still rejected for operational use. Nevertheless, the Franks suit stimulated developments that led to the present lightweight air-filled antiblackout suit, the G suit. A reexamination of the effect of complete submersion to heart level was made at the Mayo Clinic [20, 64]; the protection of about 2 G provided by this method, when compared with the 1.5-G protection of the air-filled suit, did not seem to justify further development.

Subsequently, animal studies reemphasized the potential of the water-immersion technique [43]. Rat fetuses, still without air in the lungs, survived a 10,000-G impact [44]. The protozoan *Euglena gracilis* had a 50 per cent survival of 212,000 G for 4 hr [46]. The fish *Lebistes reticulatus* survived 10,000 G for 30 sec [46]. Preoxygenated immersed mice survived 1,300 G for 60 sec [46].

In this same period Gray reexamined the conditions for human experiments [33]. He recognized distortion of the lung as a major problem to be reduced by pressurization of the lung. Using the same "Mayo tank" employed previously [20] but submerging to eye level and holding his breath from a maximum inspiration so that as his chest was compressed lung pressure increased, Gray reached 16 G_z without blackout [16, 35], a major advance in positive-G protection. This, however, required that maximum unaided lung pressurization be attained; Gray not only vigorously pressed his lips together but also taped them closed [16]. At 16 G_z air was forced from his lungs, causing vibrations of the soft palate and some slight bleeding. Mechanical aid in pressurizing the lung was therefore required. For safely pressurizing the lung at above 70 mm Hg, counterpressure about the body is required. Gray therefore designed a body-immersion capsule, nicknamed the Iron Maiden (Figure 4-3), for further acceleration studies. In this same period, Air Force studies also showed the benefits of water immersion [7].

Initial work with the water capsule was in the $-G_x$, or "eyeballs-out," position. As already stated, Collins reached -26 G_x, Webb reached -28 G_x, and Gray reached -31 G_x and held this value for 5 sec. Peak G was reached in 12.5 sec; the descent from peak also took 12.5 sec, with the centrifuge under mechanical cam control. Before the run the subject, who wore a respiration mask connected to an outside air source of variable pressure, was able to move his chest to breathe by displacing water up and down in a standpipe behind the capsule. During centrifugation, this standpipe was closed; the capsule was therefore held to con-

stant volume. Collins and Gray held their breaths during most of their runs. With Webb, lung pressure was cycled to attain oxygen transport at constant lung volume. However, the cycling pressures produced sinus pain. Before centrifugation, great care was taken to remove entrapped air bubbles from the capsule, for during the run these could expand at

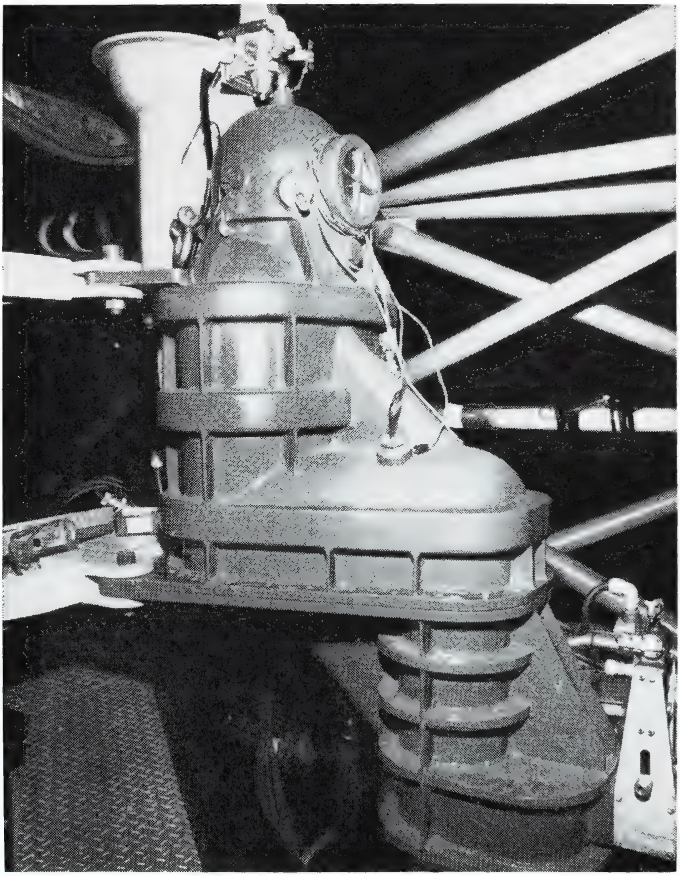


FIG. 4-3 *Iron Maiden body-immersion capsule. (Official U.S. Navy photograph.)*

the expense of lung compression. Under water during high acceleration, the arms could still be easily moved about. In some runs it was noted that the chest tended to float back and the hips forward. Moderate chest and abdominal pain was noted in some runs, but Gray had not reached any obvious tolerance end point at -31 G_x. He stopped here because the program time was finished and so that any possible delayed effects might be observed. A small amount of bloody mucus was noted on the hand-

kerchief after he blew his nose, but otherwise the experience was without apparent trauma.

A difficulty, however, is that, whereas the lung is at a constant pressure, the surrounding tissues experience a pressure gradient due to the acceleration. The tendency would be for the lung to collapse at the bottom and expand at the top along the G vector. Such expansion might force air into the tissues and blood stream. Gray had a contoured cuirass available to limit such chest distortion [16], but he did not note any feeling of such distortion during the $-G_x$ condition and so did not use the device. However, along the $+G_z$ vector, the respiratory tree is of greater length, and gradient problems would be more severe. Gray designed and built a "lower-face mask" to cover the nose and mouth and to limit expansion at the throat-to-chest level. However, it was decided to do chimpanzee experiments first in this $+G_z$ orientation. Thus far, the animal experiments that have been done [19] have involved sewing a tracheal tube in place for respiration and have not used an equivalent of the lower-face mask. Even without centrifugation, if great care is not used in eliminating air bubbles from the capsule and gas bubbles from the abdomen, pressurization of the lung leads to an overexpansion of the lung, with air being forced into the tissues. This condition is

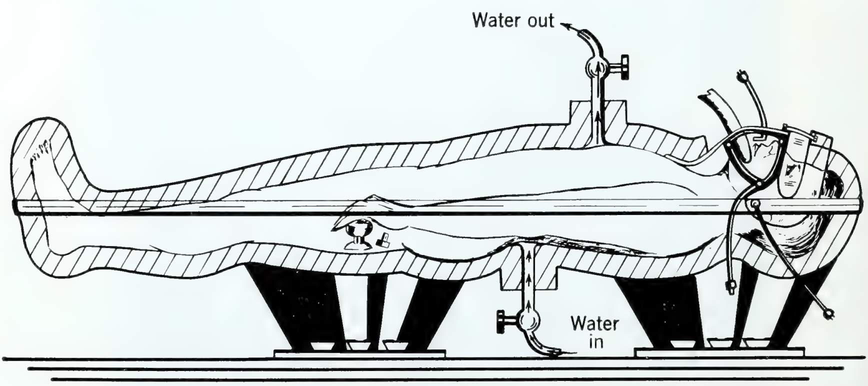


FIG. 4-4 *Minimum-weight molded-capsule concept. (Official U.S. Navy photograph.)*

worsened by adding the tissue-pressure-gradient effects of centrifugation. The work thus far has clarified some of the problems in using animals to simulate the human conditions and has provided ample material about the forcing of air into the tissues, but this work has not yet involved the use with animals of the protective equipment designed for use with man.

Lansberg has recommended water immersion in a couch inside a

capsule that is automatically positioned to be transverse to the acceleration vector [40]. Gray has recommended for minimizing weight a form-fitting water-immersion capsule that is to be entered just prior to emergency high-acceleration conditions (Figure 4-4). The problems of lung-density difference might be solved by filling the lung with fluid, perhaps a fluid able to support oxygen transport or perhaps a passive fluid with oxygen transport provided by an artificial system involving an external circulatory shunt. The latter systems would indeed involve "engineering" of the human being [49] for extending his capabilities.

CONCLUSION

This discussion has emphasized studies on the Johnsville Navy centrifuge. For a more extensive review of impact-acceleration work, see A. Martin Eiband's summary of the literature [27] and Clark and Faubert's bibliography on the biological effects of impact [15]. It is hoped that it has been sufficiently emphasized here that body-distortion tolerance rather than simply acceleration tolerance should now be studied. With the development of restraints limiting both external and internal distortions and with the possible development of procedures for safely and reversibly filling the gas spaces, it is expected that considerably higher and longer-duration acceleration than is now experimented with will be routinely experienced with comfort.

BIBLIOGRAPHY

1. Angel, Chester, et al.: Proposed Program to Develop Techniques for Simulation of Physiological Sensations of Re-entry and Landing of Boost-Glide Vehicles, *Goodyear Aircraft Corp. Rept.* TAP-8935, Dec. 21, 1959.

2. Armstrong, H. G., and J. W. Heim: The Effect of Acceleration on the Living Organism, *J. Aviation Med.,* vol. 9, pp. 199–215, 233, 1938.

3. Bailly, Le: Factors of Safety and Indexes of Static Tests, *NACA Tech. Mem.* 206, May, 1923.

4. Bauer, Louis H.: "Aviation Medicine," The Williams & Wilkins Company, Baltimore, 1926.

5. Beckman, Edward L.: Effects of Acceleration upon the Cerebral Metabolism and Cerebral Blood Flow, Phase IV: Some Observations on Negative G Developed in Acrobatics, *USN Air Developm. Center Rept.* NADC-MA-5203, April, 1952.

6. Beeding, Eli L., and Rufus Hessberg: Daisy Track Tests 271–337, Run: 4 February–19 May 1958, *Aeromed. Field Lab. Test Rept.* 8, November, 1958.

7. Bondurant, S., et al.: Human Tolerance to Some of the Accelerations Anticipated in Space Flight, *U.S. Forces Med. J.*, vol. 9, pp. 1093–1105, 1958.

8. Bührlen, L.: Versuche über die Bedeutung der Richtung beim Einwirken von Fliehkräften auf den Menschlichen Körper, *Luftfahrtmedizin*, vol. 1, pp. 307–325, 1937.

9. Cherniack, Neil S., et al.: Some Aspects of Respiratory Physiology during Forward Acceleration, *Aerospace Med.*, vol. 32, pp. 113–120, 1961.

10. Clark, Carl: Technical Film Report Med. 5-60 on Navy Centrifuge Studies for NASA Project Mercury, Program I, August 1959, *USN Air Developm. Center Rept.* NADC-MA-L6014, 1961.

11. ———: Navy Centrifuge Simulation of the National Aeronautics and Space Administration Project Mercury Vehicle, Program I, August 1959, *USN Air Developm. Center Rept. Serial* 0890, October, 1959.

12. ———: Observations of a Human Experiencing 2 G for 24 Hours, *Aerospace Med.*, vol. 31, pp. 300–301, 1960. (Abstract.)

13. ———: Some Body Displacements and Medical Effects of Lateral Accelerations during Navy Centrifuge Simulation of Ejection Capabilities from the Army A O Aircraft, *USN Air Developm. Center Rept.* NADC-MA-6044, April, 1961.

14. ——— and William Angerson: Human Acceleration Tolerance While Breathing 100% Oxygen at 5 psia Pressure, *Aerospace Med.*, vol. 32, p. 226, 1961. (Abstract.)

15. ——— and Denis Faubert: A Chronological Bibliography on the Biological Effects of Impact, *Martin Co. Eng. Rept.* 11953 and 11953 appendix, Sept. 27, 1961, Nov. 22, 1961. [See also *Rept.* 11953, revised May, 1962.]

16. ——— and R. F. Gray: A Discussion of Restraint and Protection of the Human Experiencing the Smooth and Oscillating Accelerations of Proposed Space Vehicles, *USN Air Developm. Center Rept.* NADC-MA-5914, December, 1959. [Also in P. Bergeret (ed.), "Bio-Assay Techniques for Human Centrifuges and Physiological Effects of Acceleration," AGARDograph 48, chap. 7, Pergamon Press, Inc., New York, 1961.]

17. ———, J. D. Hardy, and R. J. Crosbie: A Proposed Physiological Acceleration Terminology, with an Historical Review, in J. D. Hardy (ed.), Human Acceleration Studies, *Natl. Acad. Sci.–Natl. Research Council Publ.* 913, 1961.

18. ——— and C. H. Woodling: Centrifuge Simulation of the X-15 Research Aircraft, *USN Air Developm. Center Rept.* NADC-MA-5916, December, 1959. [See also C. H. Woodling and C. C. Clark, Studies of Pilot Control during Launching and Reentry of Space Vehicles, Utilizing the Human Centrifuge, *Inst. Aeronaut. Sci. Rept.* 59–39, 1959.]

19. Coburn, Kenneth R., Peter H. Craig, and Edward L. Beckman: Effects of Positive G on Chimpanzees Immersed in Water, *Aerospace Med.*, vol. 31, p. 226, 1961. (Abstract.)

20. Code, C. F., E. H. Wood, and E. J. Baldes: Hydrostatic Anti-blackout Protection: The Protection Afforded Man against the Effects of Positive Acceleration by Immersion in Water, *Fed. Proc.*, vol. 4, p. 15, 1945. (Motion-picture abstract.)

21. Collins, C. C., R. J. Crosbie, and R. F. Gray: Pilot Performance and Tolerance Studies of Orbital Re-entry Acceleration, Letter Report, Project TED ADC AE-1412, U.S. Naval Air Development Center, Johnsville, Pa., Sept. 19, 1958.

22. Davidson, S., B. Rose, and W. K. Stewart: Review of the Practicability of and Necessity for Anti-G Devices in the RAF with Particular Reference to the FFS Mark III, *Flying Personnel Research Comm. Rept.* 584, July, 1944.

23. Davies, H. W.: Some Medical Aspects of Flying, *Med. J. Australia,* vol. 2, pp. 1–4, July 3, 1920.

24. DeHaven, H.: Mechanical Analysis of Survivals in Falls from Heights of 50 to 150 Feet, *War Med.,* vol. 2, pp. 586–596, 1942.

25. Diringshofen, H. von: Über den Einfluss der Körpeshaltung für das Ertragen Höher Beschleunigungen, *Verhandl. Deut. Ges. Inn. Med.,* vol. 48, pp. 283–287, 1936.

26. Doolittle, James H.: Accelerations in Flight, *NACA Rept.* 203, 1925.

27. Eiband, A. Martin: Human Tolerance to Rapidly Applied Accelerations: A Summary of the Literature, *NASA Mem.* 5-19-59E, June, 1959.

28. Franks, W. R.: Test Flights with the Hydrostatic (Anti-G) Suit, *Flying Personnel Research Comm. Rept.* 301, 1941.

29. Galambos, Robert: Physiological and Psychological Testing, in James D. Hardy (ed.), Reports on Human Acceleration, *Natl. Acad. Sci.–Natl. Research Council Publ.* 901, 1961.

30. Gauer, O.: The Physiological Effects of Prolonged Acceleration, in "German Aviation Medicine in World War II," pp. 554–583, Department of the Air Force, Washington, 1950.

31. Gell, C. F., and H. N. Hunter: Physiological Investigation of Increasing Resistance to Blackout by Progressive Backward Tilting to the Supine Position, *USN Air Developm. Center Rept.* NADC-MA-5406, June 30, 1954.

32. Goldman, David, E., and Henning E. von Gierke: The Effects of Shock and Vibration on Man, *Naval Med. Research Inst. Lecture Rev. Ser.* 60–3, January, 1960. [See also their chapter, Shock and Vibration Limits of Confidence and Safety, in C. M. Harris and C. E. Crede (ed.), "Handbook of Shock and Vibration Control," McGraw-Hill Book Company, Inc., New York, 1961.]

33. Gray, R. F.: Mechanical System Suggested for G Protection, *USN Air Developm. Center Rept.* NADC-MA-5708, July, 1957.

34. ———: Water Goggles to Support the Eyes under Acceleration (patent application), U.S. Naval Air Development Center, Johnsville, Pa., 1960.

35. ——— and M. G. Webb: High G Protection, *Aerospace Med.,* vol. 32, pp. 425–430, 1961.

36. Head, Henry: The Sense of Stability and Balance in the Air, Reports of the Aviation Medical Investigation Committee, *Med. Research Council* (London) *Spec. Rept. Ser.* 28, 1919. [Reprinted in "The Medical Problems of Flying," H. M. Stationery Office, London, 1920.]

37. Hershgold, E. J.: Roentgenographic Study of Human Subjects during Transverse Acceleration, *Aerospace Med.*, vol. 31, pp. 213–219, 1960.

38. Hiatt, E. P.: Principles of Safety Monitoring of Hazardous Research on Acceleration Utilizing Human Subjects, in James D. Hardy (ed.), Reports on Human Acceleration, *Natl. Acad. Sci.–Natl. Research Council Publ.* 901, 1961.

39. Hill, L., and H. Bernard: The Influence of the Force of Gravity on the Circulation, Part II, *J. Physiol.*, vol. 21, pp. 323–352, 1897.

40. Lansberg, M. P.: "A Primer of Space Medicine," Elsevier Press, Inc., Houston, 1960.

41. Maaske, C. A., G. A. Hallenbeck, and E. E. Martin: The Effect of Taping the Body on "G" Tolerance in Man, *Natl. Res. Council Comm. Aviation Med. Rept.* 204, September, 1943.

42. MacWilliam, J. A.: Discussion of a Paper on Postural Hypotension, *Lancet*, vol. 1, p. 198, 1935.

43. Margaria, R.: La Condizione di Subgravita e la Scottrazione dall Effecto delle Accelerazione, *Riv. Med. Aeronaut.*, vol. 16, pp. 469–474, 1953.

44. ———, T. Gualtierotti, and D. Spinelli: Protection against Acceleration Forces in Animals by Immersion in Water, *Aerospace Med.*, vol. 29, pp. 433–437, 1958.

45. Marshall, G. S.: Physiological Limitations of Flying, *J. Roy. Aeronaut. Soc.*, vol. 37, pp. 389–410, 1933.

46. Morris, David P., Dietrich E. Beischer, and J. J. Zarriello: Studies on the G Tolerance of Invertebrates and Small Vertebrates While Immersed, *J. Aviation Med.*, vol. 29, pp. 438–443, 1958.

47. Oberth, H.: "Wege zur Raumschiffahrt," 3d ed., p. 108, R. Oldenbourg-Verlag, Munich, 1929.

48. Orlebar, H. H.: Discussion of the Paper "The Physiological Limitations of Flying," *J. Roy. Aeronaut. Soc.*, vol. 36, pp. 402–403, 1933.

49. Pogrund, Robert S.: Human Engineering or Engineering of the Human Being—Which? *Aerospace Med.*, vol. 32, pp. 300–315, 1961.

50. Poppen, J. R.: Report of the Physiological Effects of Sudden Change in the Speed and Direction of Airplane Flight, U.S. Navy Bureau of Aeronautics, Washington, Apr. 15, 1932.

51. ——— and C. K. Drinker: Physiological Effects and Possible Methods of Reducing Symptoms Produced by Rapid Changes in Speed and Direction of Airplanes as Measured in Actual Flight, *J. Appl. Physiol.*, vol. 3, pp. 204–215, 1950.

52. Smedal, Harold A., Brent Y. Creer, and Rodney C. Wingrove: Physiological Effects of Acceleration Observed during a Centrifuge Study of Pilot Performance, *NASA Tech. Note* D-345, December, 1960.

53. Stapp, John P.: Effect of Linear Acceleration, in Harry Armstrong (ed.), "Aerospace Medicine," pp. 263–268, The Williams & Wilkins Company, Baltimore, 1961.

54. Swearingen, John, et al.: Human Voluntary Tolerance to Vertical Impact, *Aerospace Med.*, vol. 31, pp. 989–998, 1960.

55. Thurston, A. P.: Discussion of the Paper of G. T. R. Hill, Advances

in Structural Design, the Third Anglo-American Aeronautical Conference, Sepember, 1951, Royal Aeronautical Society, London, 1952.

56. Turning Somersaults with an Aeroplane: The Remarkable Exploit of Adolphe Pégoud, *Sci. American,* vol. 109, p. 240, 1913.

57. Vykukal, H. C., G. W. Stinnett, and R. P. Gallant (NASA Ames Research Laboratory): Performance of an Interchangeable Mobile-Pilot-Restraint-System Designed for Use in a Moderately High Acceleration Field, paper presented at the Aerospace Medical Association Meeting, Chicago, Apr. 26, 1961. [See also George Alexander, NASA Develops High-Mobility Space Suit, *Aviation Week,* vol. 75, no. 1, pp. 57–59, July 3, 1961.]

58. Weaver, John, et al.: Direct Casting of a Human in Rigid Polyurethane Foam for Use as a Restraint System in High Acceleration Environment, *USN Air Developm. Center Rept.* 6146, June, 1962.

59. White, Stanley C.: Review of Medical Systems for MR-3 Flight, *Proc. Conf. Results First U.S. Manned Suborbital Space Flight,* Government Printing Office, Washington, June, 1961.

60. Wiesehöfer, H.: Über Flugversuchezur Frage der Erträglichkeit Höher Beschleunigungen bei Liegender Unterbringung der Flugzeuginsassen, *Luftfahrtmedizin,* vol. 4, pp. 145–155, 1940.

61. Wood, E. H.: Use of the Valsalva Maneuver to Increase Man's Tolerance to Positive Acceleration, *Fed. Proc.,* vol. 6, p. 229, 1947. (Abstract.)

62. ———, D. M. Clark, and E. H. Lambert: An Analysis of Factors Involved in the Protection Afforded Man by Pneumatic Anti-blackout Suits, *Fed. Proc.,* vol. 4, p. 79, 1945.

63. ———, C. F. Code, and E. J. Baldes: Protection against the Effects of Acceleration Afforded the Human by Assumption of the Prone Position, *Natl. Research Council Comm. Aviation Med. Rept.* 158, March, 1943.

64. ———, ———, and ———: The Protection against the Effects of Acceleration Afforded the Human by Immersion in Water and by a Water Filled Suit (the FFS), Wright Field, Dayton, Ohio, August, 1943.

65. ——— and G. H. Hallenbeck: Voluntary (Self Protective) Maneuvers Which Can Be Used to Increase Man's Tolerance to Positive Acceleration (with introduction by C. F. Code), *Fed. Proc.,* vol. 4, pp. 78–79, 1945.

66. ——— and E. H. Lambert: Some Factors Which Influence the Protection Afforded by Pneumatic Anti-G Suits, *J. Aviation Med.,* vol. 23, pp. 218–228, 1952.

67. ———, ———, and C. F. Code: Do Permanent Effects Result from Repeated Blackouts Caused by Positive Acceleration? *J. Aviation Med.,* vol. 18, pp. 431–432, 1947.

Chapter 5

HUMAN RESPONSE TO VIBRATION

Edward B. Magid and Rolf R. Coermann†*

OSCILLATORY MECHANICAL ENVIRONMENTAL FORCES EXTRINSICALLY APPLIED to man have been constantly increasing in magnitude and complexity. The etiology of these forces covers a wide range of sources, from man's own movements of walking, running, and jumping to oscillations resulting from mechanical and electrical devices including hand-operated tools, heavy industrial machinery, and vehicles of all types. These oscillations include a wide spectrum of frequencies, ranging from repetitive single impacts to vibrations of extremely high frequencies involving the entire body and/or specific body regions.

One of the earliest physiological reports on vibration and man was made on the syndrome called "white fingers," or "secondary Raynaud's disease," in pneumatic-hammer operators [13]. This phenomenon is characterized by symmetrical blanching of the distal portions of the fingers and sensations of numbness and paresthesias including tingling, numbness, and burning [11]. Blanching is followed by cyanosis and later by redness. The syndrome is precipitated by cold or psychic stimuli. It was found that the frequency of these hammers usually ranged from 60 to 2,000 cps [12]. When the tools produced frequencies below 250 cps, there was a tendency to produce skeletal abnormalities, particularly of the articulating surfaces, as seen in Figure 5-1. Above 250 cps, abnormalities in local nervous and vascular tissue tended to occur.

Abnormalities resulting from whole-body vibrations have also been observed. Investigators have surveyed the incidence of physical disorders in drivers of trucks, tractors, motorcycles, and other vehicles or machinery in which appreciable vibrations and jolts occur. These workers exhibited a high incidence of osteoarthritis, traumatic fibrositis, herniated disks, coccygodynia, lumbosacral pain [8], abdominal pain, and intestinal

* Department of Physiology, Chicago Medical School, and Department of Medicine, Cook County Hospital, Chicago, Ill. Formerly of Aerospace Medical Laboratory, U.S. Air Force, Wright-Patterson Air Force Base, Ohio.

† Vibration and Impact Section, Bioacoustics Branch, Aerospace Medical Laboratory, U.S. Air Force, Wright-Patterson Air Force Base, Ohio.

This chapter represents the view of the authors, not necessarily the views or sanction of the U.S. Air Force.

disorders [22]. Visceroptosis was a common cause of vague abdominal distress. Figure 5-1 diagrammatically illustrates the pathology resulting from (1) the vibratory stimuli, (2) the direction of the force, and (3) the point of excitation.

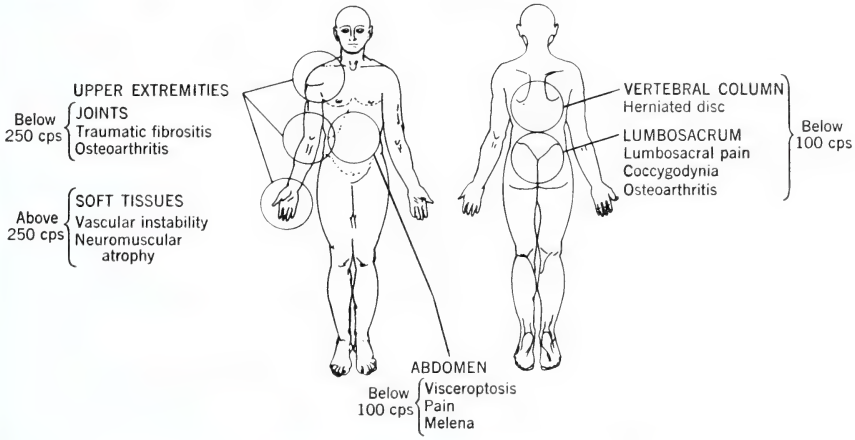


FIG. 5-1 *Common regional pathology resulting from occupational vibratory environments.*

During the past thirty years investigators from many disciplines have studied the biological effects of vibration in the laboratory. Pathological studies have been performed on mice, rats, cats, dogs, and monkeys. The lesion most commonly found in most body regions was that of hemorrhage. Petechial or frank hemorrhages were found in the lungs, myocardium, gastrointestinal tract, kidneys, and central nervous system [7]. Histological studies showed changes in the architecture of vascular connective tissue. These changes demonstrated the vulnerability of the ubiquitous vasculature to the destructiveness of whole-body vibrations. Recent work has suggested that direct mechanical excitation is only part of the pathogenesis of the trauma and that other factors such as hormonal imbalance resulting from vibration also play a part.

Many physiological laboratory observations have also been made of man. Among the early workers in this field, Coermann utilized human subjects in studies of frequencies above 20 cps [1]. Increases in pulse rate, respiratory rate, and blood pressure were found to occur at certain frequencies. In addition, the patellar reflex was found to be inhibited. Other workers found these parameters to show little or no change or actual decreases; it should be noted, however, that their investigations utilized different frequencies and experimental procedures.

Vasomotor instability similar to the condition found in man was produced in rats by standing animals on a vibrating platform. It was

ascertained that mechanical stimuli such as individual impacts or vibrations generally cause vasoconstriction on the arteriole or capillary levels [10]. Dermographia and acrocyanosis as a result of vibrations to the extremities have also been reported in experimental animals.

Water and electrolyte metabolism have shown changes. Animals subjected to vibration have shown an altered water balance as well as an increase in sodium and potassium turnover. Hormones effecting water and electrolyte balance have also shown changes. For example, increases in adrenal cortical secretions and changes of the estrus cycle in the rat have been demonstrated [19].

The effects of vibration on the performance of man and experimental animals have been the subject of much investigation. Little decrement in tapping, speech, grip strength, or mirror tracing has been reported. However, reduced foot-pressure consistency in human beings, decreased jumping ability in monkeys, and increase in time for rats to go through a maze have been reported [18].

The preceding observations and studies utilized various animal species, including man. These animals varied in size, anatomical configuration, and organ-tissue characteristics. They also showed various differences in physiology and biochemistry. Experimental frequencies, amplitudes, couplings to the extrinsic forces, and directions of force varied greatly and in many cases were not recorded in the reports. Certainly these studies have demonstrated the types of physiological changes that are caused by vibrations. In order to understand the pathological and physiological changes that occur and to be able to compare these changes among the various species and develop concepts that may be practically utilized, it is necessary to understand the body dynamics of the experimental animal and the transmission of forces involved.

PHYSICAL RESPONSE TO VIBRATION

The mechanical response of the human body to vibratory energy may be roughly divided into low-, middle-, and high-frequency ranges, which exhibit different characteristics [8]. Up to approximately 100 cps the body acts as a complex system of masses, elasticities, and dampers having lumped parameters. Large organ-tissue complexes having viscoelastic characteristics are coupled to a skeletal framework consisting of relatively rigid bony components that are tightly bound together. The frequencies varying from approximately 100 to 100,000 cps are characterized predominantly by wave propagations of vibratory energy, including shear waves, surface waves, and compressional waves; all three are greatly influenced by boundaries and structural configuration. From 100,000 cps

up to the range of megacycles, compression waves propagating in a beam-like manner predominate.

Manned high-performance jet aircraft, rocket-propelled space vehicles, and escape systems from supersonic craft involve low-frequency high-amplitude vibrations. In this frequency range, particularly from 0.5 to 20 cps, man absorbs most of the vibratory energy from his environment because his built-in vibration-isolation capacity is least effective in this range. This phenomenon can be explained by the several apparent resonances of body structure that occur within this frequency range. Therefore, further discussion is restricted to whole-body vibrations within the low-frequency range, i.e., up to 20 cps.

The human body exposed to vibrations and buffeting in this low-frequency range responds in the same manner that an inorganic system with lumped parameters does. Because of the extremely complicated structure of the body, the responses for frequency, magnitude, kind of excitation, and duration of stresses differ from those of an inorganic system. However, the principal difference between an inorganic system and a living organic system is that the inorganic system does not change as long as the strain does not exceed the elastic limits of any of its parts. In an organic system, however, any strain will act as a stimulus affecting tissues, including nervous receptors, and the accumulation of these stimuli might produce irreversible changes in the organism even though soft tissue or bone might fail to show overt morphological changes. The many control systems of the body are relatively sensitive to long-term overload. An investigation of the effects of dynamic forces on the human body must consider not only the damage to soft tissues and bones but also all observable physiological and psychological effects.

Impedance Studies

One of the most revealing ways of studying the mechanical properties of the whole body is the measurement of the body's mechanical impedance [3]. This technique aids in describing mechanical energy transmission to a complex mechanical system. With the application of an alternating force to the body, the impedance is defined as the ratio of the transmitted force to the velocity of the point at which the force is applied. The transferred force, the velocity of the shake table, and the phase between force and velocity are measured at various frequencies.

If the human being had in fact been a solid mass, he would have reacted in a linear fashion as indicated by the solid line, $m\omega$ (mass $\times$ angular acceleration), in Figure 5-2, which shows the mechanical impedance of a solid mass weighing 150 lb.[1] Instead the human being

[1] This weight was selected to represent the mean weight of an adult male.

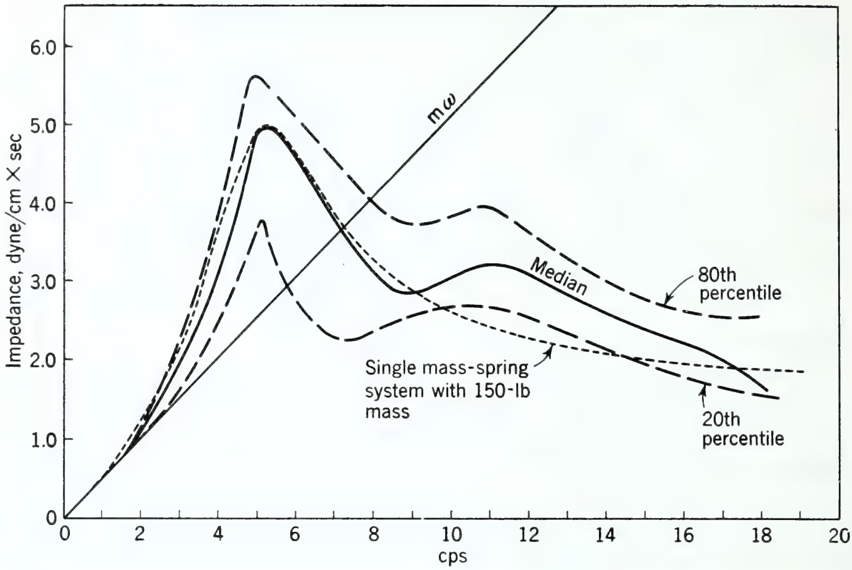


FIG. 5-2 *The median and the twentieth and eightieth percentiles of the impedance of five different subjects as compared with the impedance of a single mass-spring system and a pure mass system (mω). (After R. R. Coermann et al., The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, Aerospace Med., vol. 31, pp. 443–455, June, 1960. By permission of the publishers.)*

reacts similarly to a complex mass-spring system with two resonant peaks at 5 and 11 cps as shown in the figure. From both resonant peaks the effective masses, elasticities, and damping factors for these frequencies can be calculated. A single mass-spring system with a mass of 150 lb shows only one peak. Up to 9 cps the body behaves as a simple mass-spring system with a viscous damper. Between 9 and 15 cps the second resonance dominates.

The impedance curve changes considerably if either the posture of the subject is varied or the pelvis of the subject is inclosed in a semirigid envelope (Figure 5-3). The relaxation of all skeletal muscles and the stiffening of the pelvic area cause a marked decrease of the impedance, particularly at the higher frequencies.

The same kinds of changes are evident when the phase between the translated force and the velocity of the shake table (Figure 5-4) are taken into account. Whereas the phase of the unprotected subject turns at frequencies to —90°, indicating a change to a pure elasticity phase, the phase of the subject in a semirigid envelope stays approximately constant after the first resonant peak, suggesting that the body acts as a pure mass.

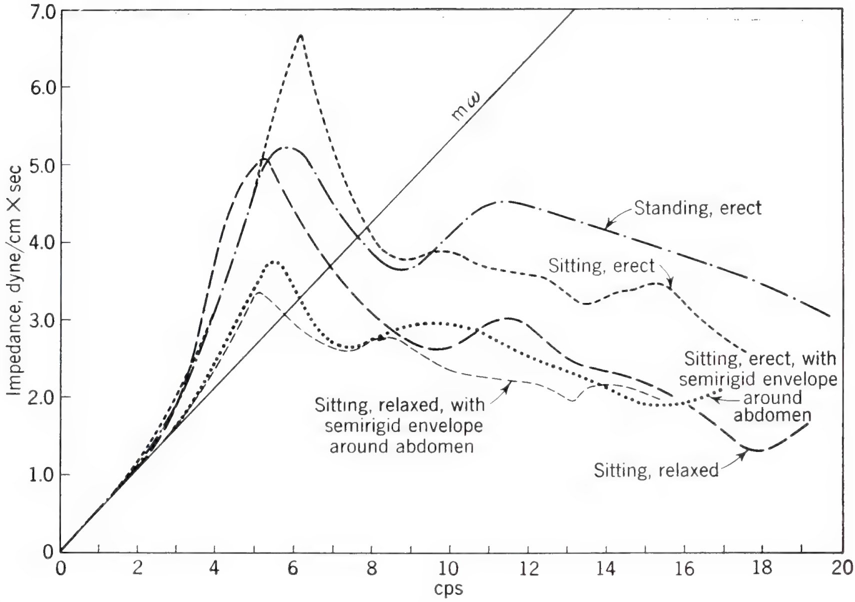


FIG. 5-3 *The mechanical impedance of one subject sitting and standing in varied postures and with varied supports. (After R. R. Coermann et al., The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, Aerospace Med., vol. 31, pp. 443–455, June, 1960. By permission of the publishers.)*

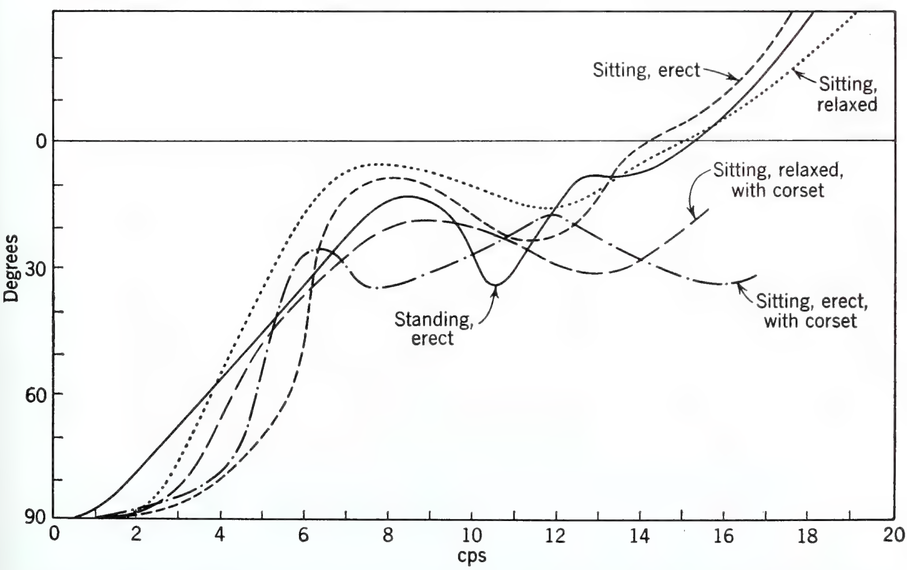


FIG. 5-4 *Phase shift of force vs. velocity as measured on one subject in varied postures. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

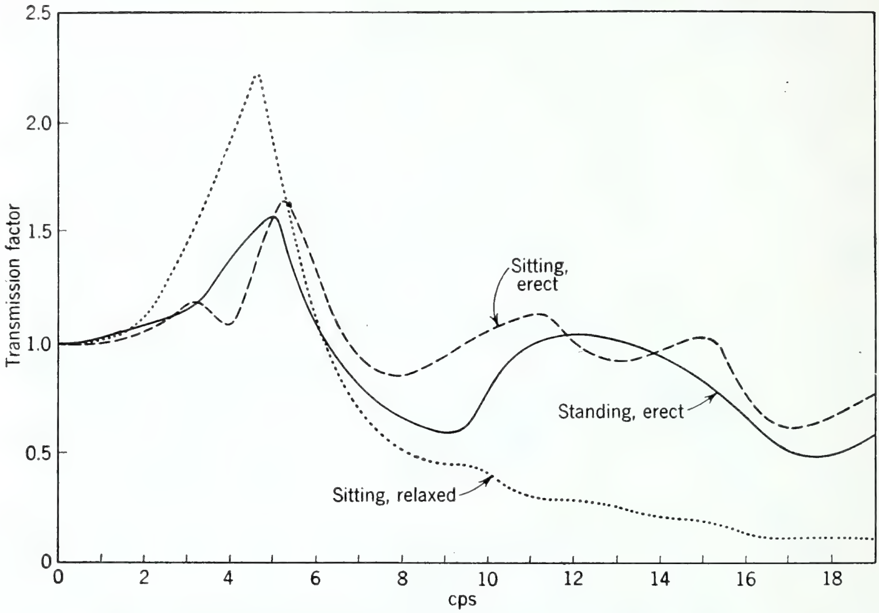


FIG. 5-5 *The transmission of vibrations from the shake table to the head of one subject with varied postures. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

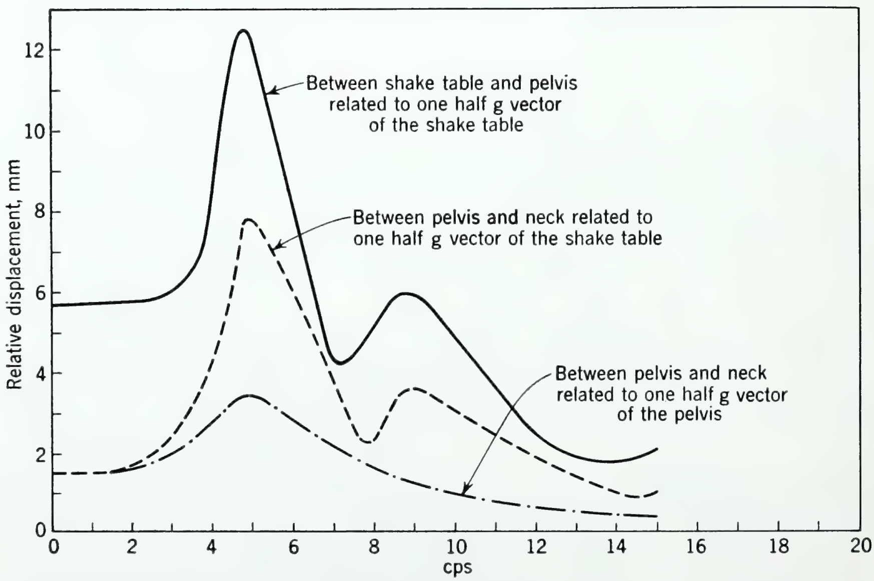


FIG. 5-6 *Relative body displacements of one subject sitting erect during vertical vibration. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

Transmission Studies

Measurement of transmission factors for vibrations transmitted to the head is another technique for gaining insight into the mechanical properties of the human body. The transmission factor is defined as the ratio of the movements of the head to the movements of the shake table. Resonant peaks similar to the impedance curves also appear here, as indicated in Figure 5-5. With relaxed posture the peak at 5 cps rises markedly, although it drops at higher frequencies.

The relative displacements between the pelvis and the shake table and between the pelvis and the neck have been measured in order to determine the respective resonances of the spine and pelvis. The pelvis had two resonances in the observed frequency range of 5 to 9 cps; the spine had only one resonant peak, which was at approximately 5 cps, as indicated by the dot-dash line in Figure 5-6.

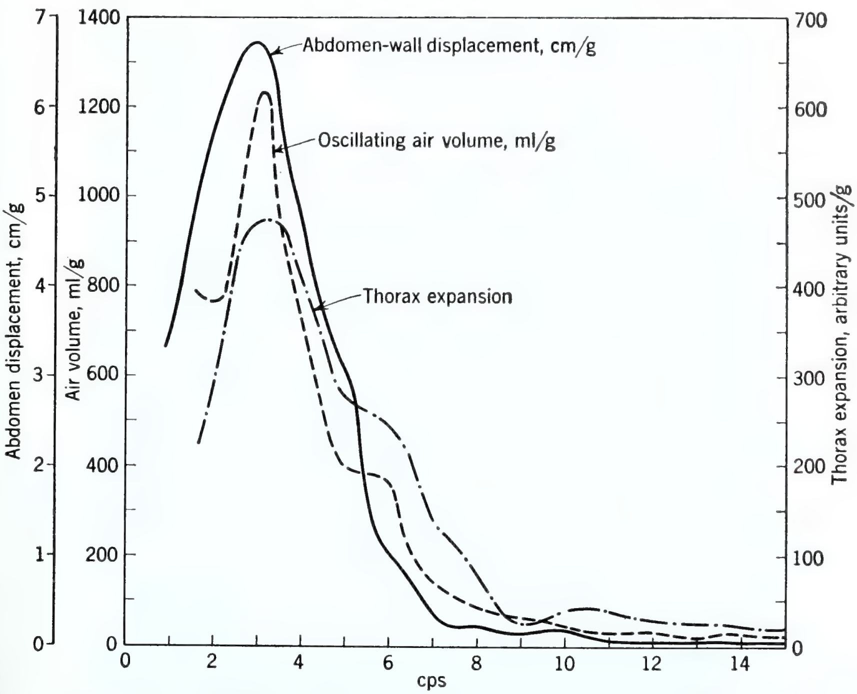


FIG. 5-7 *Abdomen-wall displacement, thorax expansion, and oscillating air volume of a human subject (supine position, shaken horizontally). (After R. R. Coermann et al., The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, Aerospace Med., vol. 31, pp. 443–455, June, 1960. By permission of the publishers.)*

One of the most important systems limiting human tolerance to vibration is the thoracoabdominal system. This system consists of the mass of the abdominal wall, viscera, the diaphragm, portions of the lungs, portions of the thoracic cage, and the inherent elasticities of the abdominal wall, diaphragm, and thorax. Extreme displacement of these body parts may produce pain or even damage to tissues of the inner organs. To determine the natural frequencies of this system, the acceleration of the abdominal wall was measured at two different points. The alternating expansion of the thorax and the volume of air entering and leaving the mouth with open glottis were also measured. Figure 5-7 shows the main resonance at approximately 3 cps, with slight peaks between 5 and 6 cps and at approximately 11 cps. The last two peaks are the resonances of the pelvis. The inner organs present a resonance at approximately 3 cps, with a relatively small amount of damping. The same

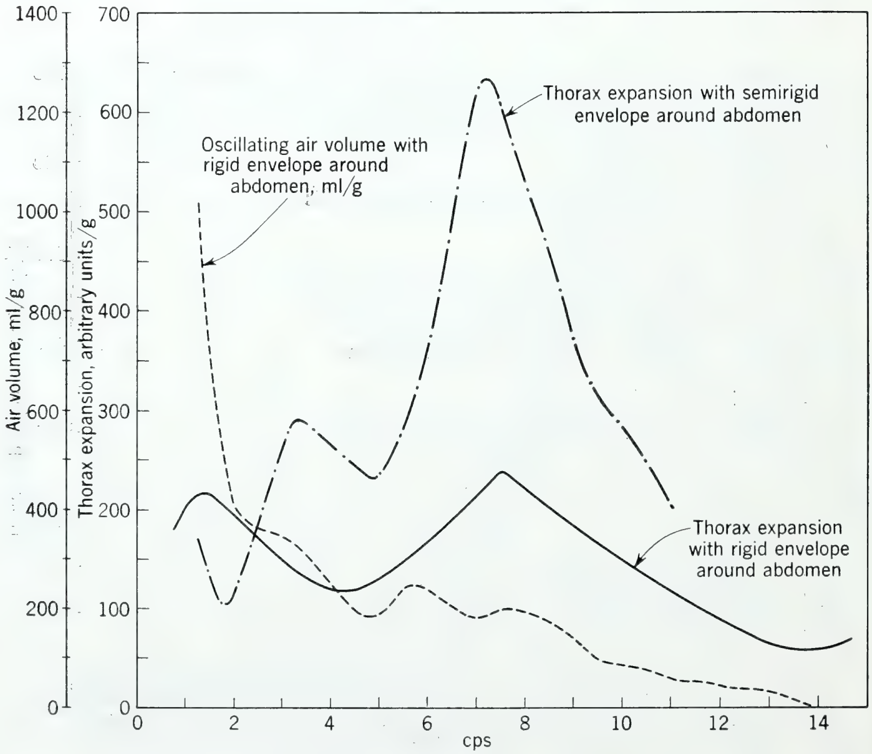


FIG. 5-8 *Thorax expansion and oscillating air volume of a human subject with envelope around abdomen (supine position, shaken horizontally). (After R. R. Coermann et al., The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, Aerospace Med., vol. 31, pp. 443–455, June, 1960. By permission of the publishers.)*

natural frequency can be calculated by taking roentgenograms of the same subjects in upright, supine, and upside-down positions and determining the static displacements of the diaphragm.

When semirigid or rigid envelopes are applied around the abdomen, the movement of the thorax and the alternating air flow changes considerably, as shown in Figure 5-8. With the semirigid envelope (a corset), the 3-cps resonance was suppressed, but, because of the elasticity of the envelope and abdominal wall, a resonant peak between 7 and 8 cps appeared. The rigid envelope, consisting of a plaster cage around the abdomen, also suppressed the 3-cps resonant peak, shifted the lowest resonance to 1.5 cps, and caused a resonant peak between 7 and 8 cps. The increase of air volume at the lower frequencies is probably due to the high compliance of the lung tissue when the thorax is passively expanded by the mass of the abdomen and no air pressure is in the lungs. This effect was further investigated by using a partial-pressure suit. The

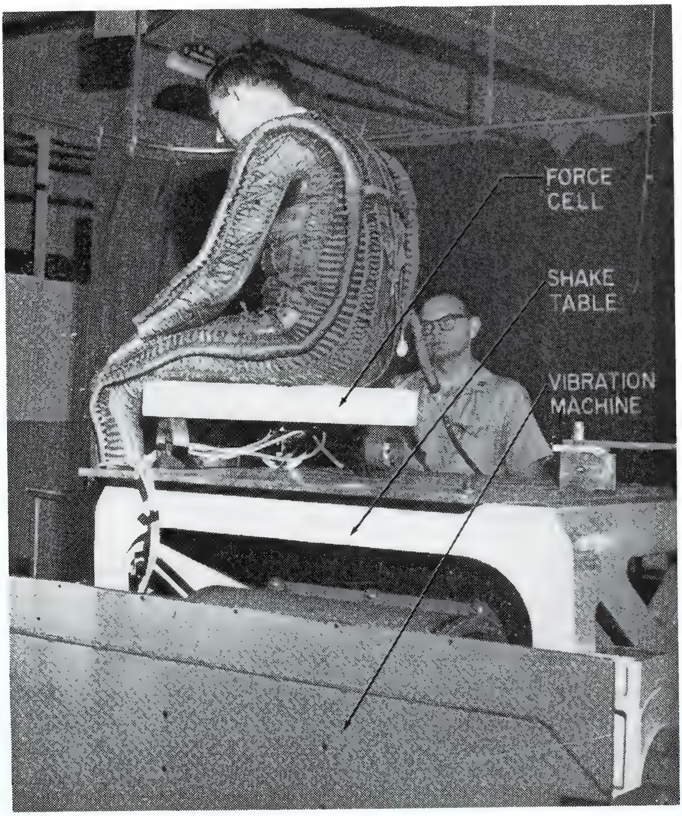


FIG. 5-9 *Partial-pressure suit, force-cell table, and vertical shake table used to study the effect of low-frequency vibration on mechanical impedance of the body.*

subject sat on a force-cell table connected to a vertical shake table, as pictured in Figure 5-9. The peripheral tubes of the suit were pressurized to 5 psi. Pressurization increases the impedance at the resonant peaks and, apparently, increases the elasticity of the body. Protection is afforded under conditions of rapid decompression, but under certain conditions of vibration the same gear does not give protection and may actually add to the deleterious effects of these vibratory forces, as shown in Figure 5-10.

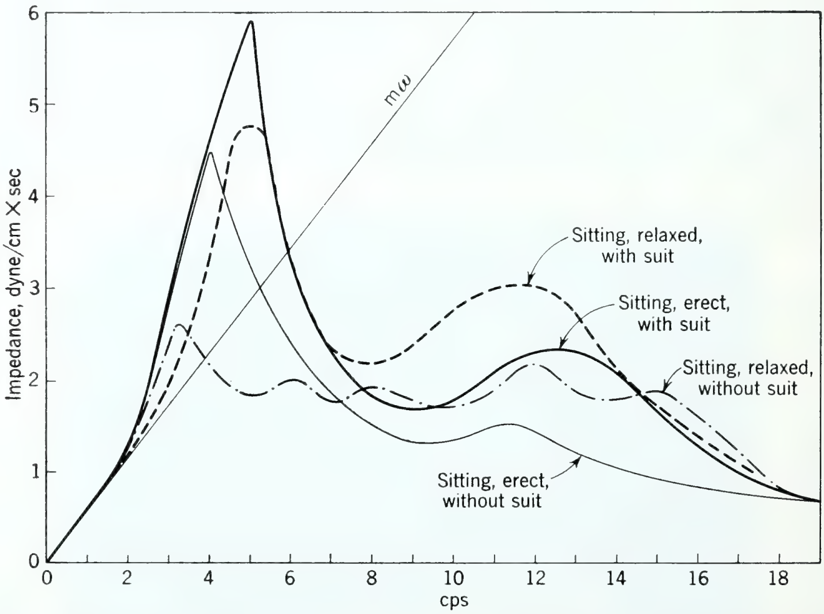


FIG 5-10 *Effect of partial-pressure suit on mechanical impedance of one subject with change in posture. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

With the data provided in this study, a model of the mechanical characteristics of the human body can be designated, as shown in Figure 5-11. The natural frequency of the head and neck system has been determined by Dieckmann to be approximately 20 cps [4]. The upper torso is supported by the spine, and the spinal column presents its lowest natural frequency at 5 cps. Attached to the upper torso are the arm-shoulder system and the thoracoabdominal system, both with a natural frequency of 3 cps. The pelvis has two natural frequencies, 5 cps and 9 cps.

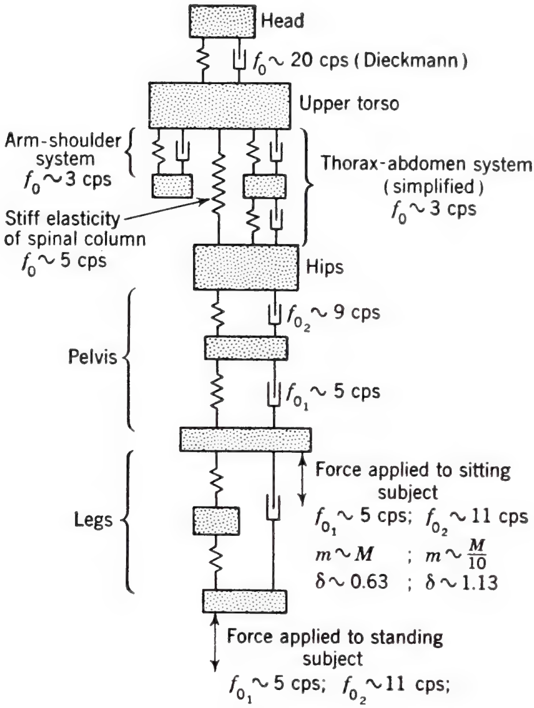


FIG. 5-11 *Mechanical analog of the human body. (After R. R. Coermann et al., The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, Aerospace Med., vol. 31, pp. 443–455, June, 1960. By permission of the publishers.)*

SUBJECTIVE RESPONSE

The studies discussed thus far have indicated some of the mechanical properties of the body during vibration. It is apparent that physiological manifestations due to whole-body vibration are dependent upon alternating displacements of organ systems and their supporting structures. The magnitude of displacement is dependent upon anatomical configuration, direction of transmitted forces, and the dynamic properties of individual body parts.

Various subjective responses, from the threshold of perception to the feeling of pain, have been defined and used to classify vibratory sensations. In reviewing the results of several investigations concerned with the willingness of subjects to tolerate various levels of vertical vibrations, Goldman has shown that the variability among studies is partially

a result of the different subject positions assumed on the vibrating platform [9].

Short-time Tolerance Study

Studies of the mechanical properties of the body define the critical frequency range but not the tolerable acceleration for each frequency. A study for determining the short-time human tolerance to sinusoidal vibration was conducted with a panel of 10 healthy males as subjects [20]. Each subject assumed the sitting position simulating that of a passenger in a commercial or military conveyance. He was restrained with a lap belt and shoulder straps so that the forces could be transmitted to the body with as much uniformity as possible. The subject sat with his coccyx pressed firmly against the back of a rigid seat, and the lap belt and shoulder straps were tightened so that movement in the vertical or

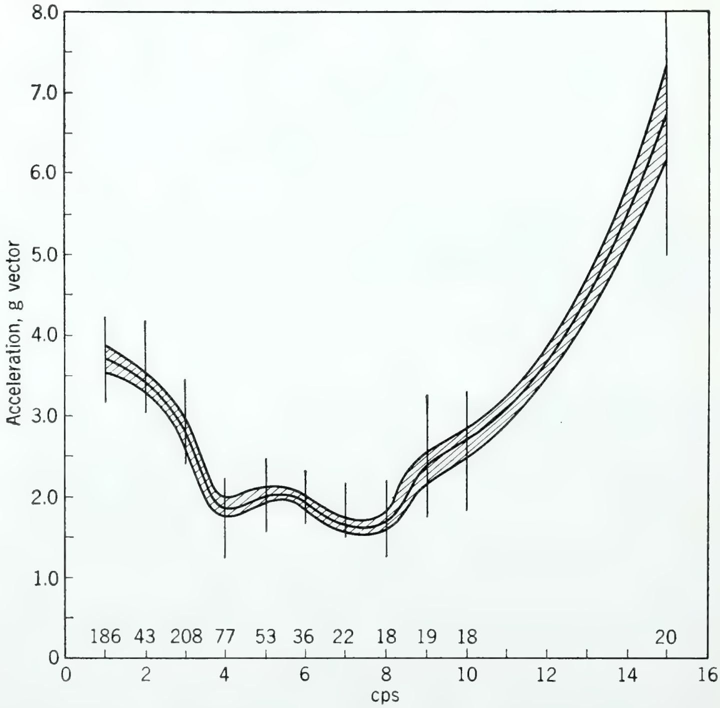


FIG. 5-12 *Arithmetic mean, standard deviation, and maximum deviation of tolerance values. The arithmetic mean of exposure time in seconds is indicated above the frequency scale for each test frequency. The vertical lines indicate the range of tolerance responses. (After E. B. Magid, R. R. Coermann, and G. H. Ziegenruecker, Human Tolerance to Whole Body Sinusoidal Vibration, Aerospace Med., vol. 31, pp. 915–924, November, 1960. By permission of the publishers.)*

horizontal direction was at a minimum. Care was taken to avoid impairing respiratory movements and changing the mechanical characteristics of the abdomen and its contents.

With the subject seated as described, the frequency was preset, and, starting from zero, the peak-to-peak amplitude was increased at 0.75 mm/sec until the subject stopped the run. Each subject was told that this study was an attempt to define the limits of sinusoidal acceleration that an individual would be willing to undergo before he felt that actual body harm would occur. At the end of each run, the reasons that each subject gave for aborting the run were recorded. The results of the experiment are presented in Figure 5-12. Measurements were taken at each frequency from 0 to 10 cps and then at 15 cps. The data at 10 to 15 cps as shown on the graph are extrapolated from other studies. The vertical lines indicate the range of tolerance responses at each measured frequency for the 10 subjects. Minimal tolerance occurs between 4 and 8 cps at accelerations between $1\frac{1}{2}$ and 2 g vector.

Table 5-1 lists reasons given by the subjects for stopping the run.

TABLE 5-1 *Number of Sensations Reported*

Cps	Sensations						
	Abdominal pain	Chest pain	Testicular pain	Head symptoms	Dyspnea	Anxiety	General discomfort
1	...	...	...	...	8	...	3
2	...	...	...	...	8	...	4
3	2	2	...	...	5	1	5
4	2	2	...	2	3	2	5
5	...	4	...	...	...	1	6
6	3	4	...	1	...	...	4
7	2	5	1	1	...	...	1
8	1	4	...	1	...	2	3
9	2	4	...	...	1	...	5
10	1	1	3	2	...	1	10
15	...	...	...	...	...	...	8

SOURCE: E. B. Magid, R. R. Coermann, and G. H. Ziegenruecker, Human Tolerance to Whole Body Sinusoidal Vibration, *Aerospace Med.*, vol. 31, pp. 915–924, November, 1960.

There are seven sensations, four of which are pain related to specific regions, the head, thorax, abdomen, and testicles. The remaining three sensations are void of pain, viz., dyspnea referable to the respiratory system, general discomfort referable to a generalized stimulation of the

sensorium, and anxiety. Abdominal and chest pains occur at 3 to 10 cps, dyspnea at 1 to 3 cps, and general discomfort over the entire range.

All pain sensations were described as having a subtle onset, being dull in nature and of low intensity, and increasing with crescendo-like characteristics as the ride continued. During the test run all subjects experienced the sensation of displacement of all body areas; this is referred to as general discomfort. This sensation was considered as the limiting factor, alone or in combination with other symptoms, only when the intensity of the displacements became so severe that the subject felt that bodily harm might occur. The abdominal pain had a slow onset, was dull and aching in nature, and gradually increased in intensity as the run continued. The pain distribution was at or below the umbilicus and often radiated to the right lower quadrant. The chest pain had a distribution resembling that of the pain experienced in coronary heart disease, a dull, aching pain that increased in intensity as the ride continued. The pain occurred at the substernal or precordial area. Testicular pain was described as pain of the groin and colicky in nature. The head sensation was described as a dull, aching pain of low intensity or a full, congested feeling, as if the head were expanding.

The short-time tolerance study served two important purposes: it defined for further study the tolerance limits of acceleration within this frequency range, and it suggested several important avenues of physiological investigation.

One- and Three-minute Tolerance Studies

After the determination of the short-time tolerance, the next step was to determine a relatively long-time tolerance to sinusoidal vibration [15]. One- and three-minute periods were chosen for the frequency ranges of 1 to 20 cps and 1 to 10 cps, respectively. It had been fairly easy to determine short-time tolerance because the shake table could be set for a given frequency and, starting from zero, the acceleration could be increased at a constant rate until the tolerance limit was reached. The accelerations for the long-time tolerance study were much more difficult to ascertain in advance, for it was impossible to predetermine the exact acceleration at which the subject would reach the tolerance limit. However, the acceleration for each frequency was estimated prior to the running of the experiment.

Figure 5-13 shows the 1- and 3-min tolerance curves, together with the short-time tolerance curve. As has been indicated, the accelerations were estimated and predetermined for each frequency. The subjects reported that they had had to exert great effort to finish the run. At the

completion of the run, the subjects estimated how much more acceleration they could have endured. The difference between the actual and the estimated accelerations was approximately ½ g at 1 to 3 cps and 8 to 20 cps and ¼ g at 4, 5, 6, and 7 cps.

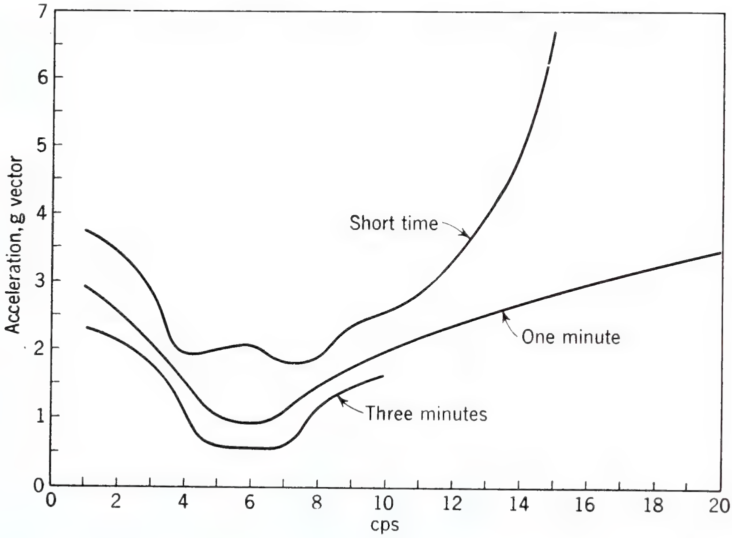


FIG. 5-13 *Human whole-body tolerance to sinusoidal vibration. (After E. B. Magid, R. R. Coermann, and G. H. Ziegenruecker, Human Tolerance to Whole Body Sinusoidal Vibration, Aerospace Med., vol. 31, pp. 915–924, November, 1960. By permission of the publishers.)*

The total number of subjective responses for all sensations at each frequency (1 and 2 cps are extrapolated data) is shown as a smoothed curve in Figure 5-14. Also presented in this figure are the responses obtained at each frequency (1 and 2 cps are extrapolated data) for the thorax, skeletal musculature, and abdomen. Each bar represents the percentage of the total number of possible responses for a body region (thorax = 160 possible responses, skeletal musculature = 120 possible responses, abdomen = 80 possible responses). It can be seen that at 6 and 7 cps there is a striking increase in the number of responses referable to the thorax, as well as an increase in the number of responses for the skeletal musculature and the abdomen. Figure 5-15 shows the response due only to pain in the back, chest, or abdomen and does not include the other symptomatology. The smoothed curve represents the total response; each bar represents the percentage of the total possible responses for a body region (lumbosacral = 40 possible responses,

chest = 40 possible responses, abdomen = 40 possible responses). Figures 5-14 and 5-15 indicate that at the frequencies studied the thorax, abdomen, and skeletal musculature are all involved in producing perceptible sensations and that the sensations limiting tolerance usually include pain.

Clearly, certain tissue-organ systems determine the tolerance limit for each frequency. Thus, if the accelerations or the time period for the 1- and 3-min tolerance levels were increased, it would be expected that these systems could be damaged and that other organ complexes would

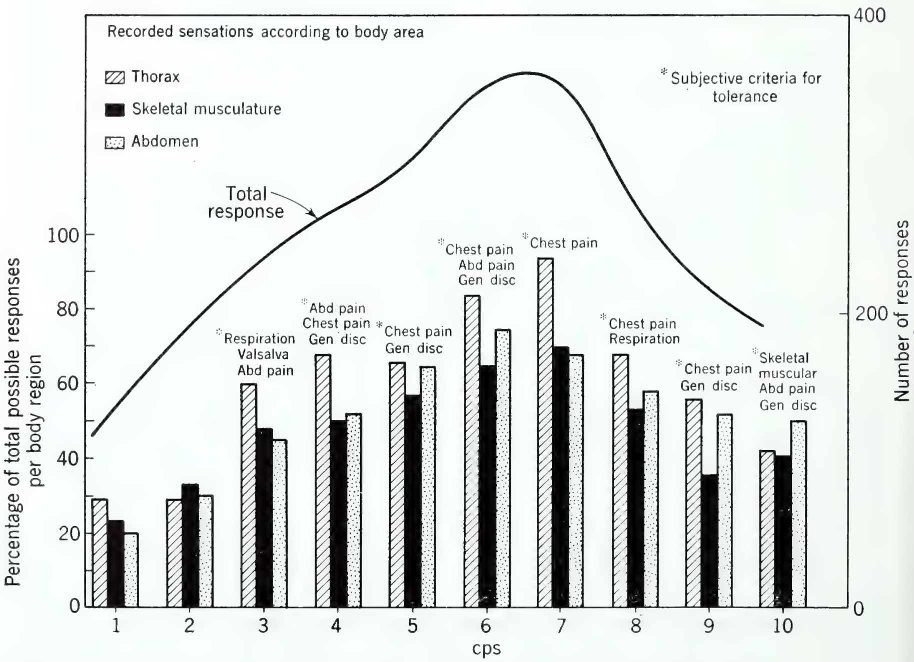


FIG. 5-14 *Subjective response, results of 10 subjects. Responses include:*

Respiration *Dyspnea* *Valsalva* *Chest pain*	*Thorax*
Skeletal muscles *Back pain* *General discomfort*	*Skeletal musculature*
Abdominal contraction *Abdominal pain*	*Abdomen*

(After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)

reach tissue tolerance limits that would lead to even more pathological changes.

The results of these studies indicate that the resonance of the thoracoabdominal system limits tolerance at 4 to 8 cps. Contained within

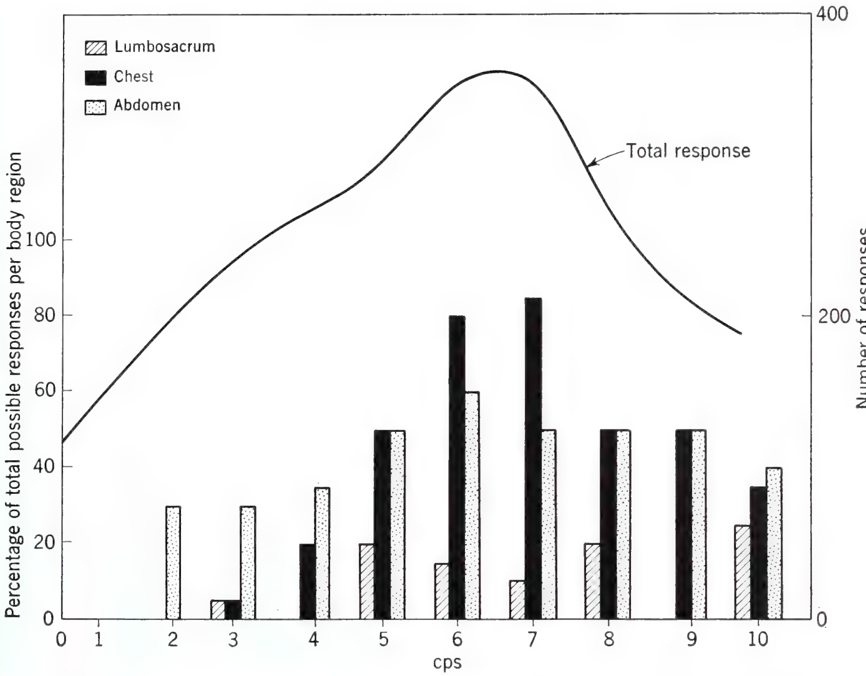


FIG. 5-15 *Subjective response due to pain, results of 10 subjects. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

the two compartments of this system are organ systems vital to body integrity. An important characteristic of the visceral organs is the large degree of freedom necessary for maintaining their physiological functions. The organs' inherent mobility is reflected in their low resonant frequencies. As previously stated, the resonant peaks for the abdomen and thorax have been found to be about 3 cps. Therefore, if it is assumed that sensations are related to resonance, it would be expected that, for a constant acceleration, sensations would change with change in frequency. One phase of the study was the investigation of the frequency range wherein resonances of the thoracoabdominal system lie [14]. With the subject seated and strapped in as in the study of short-time tolerance, an acceleration of ½ g was held constant for each frequency during a 3-min vibratory period; this acceleration was just below that for the 3-min

tolerance at 5 and 7 cps. The total number of responses for each frequency were added together and were plotted as shown in Figure 5-16. The data

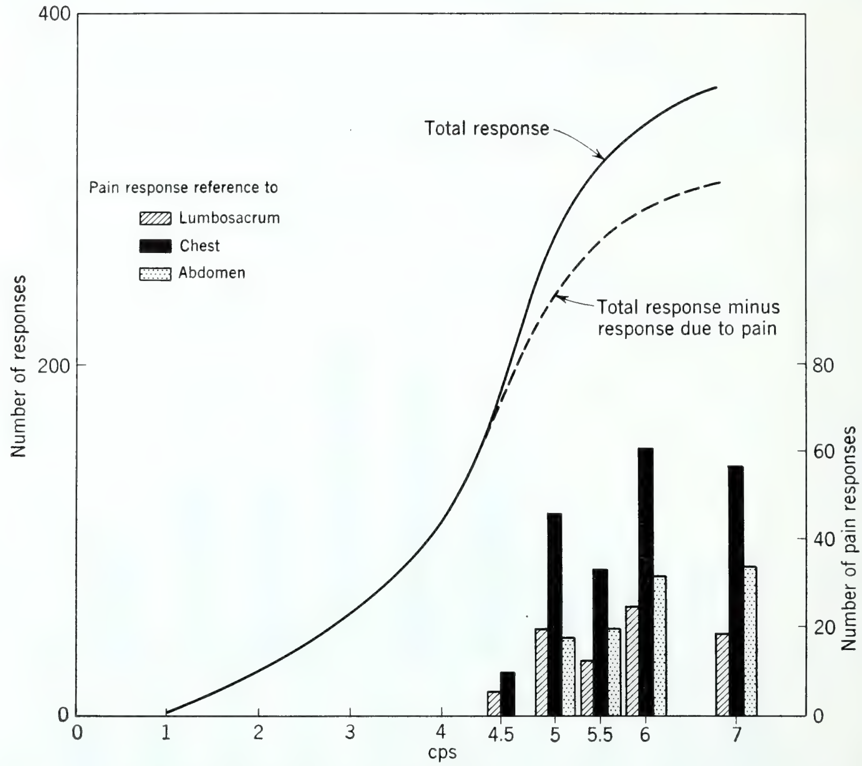


FIG. 5-16 *Subjective response to whole-body vibrations at constant acceleration of ½ g vector. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

of this part of the study, in combination with data from the previous investigation, demonstrated that for a constant acceleration of ½ g the "critical" frequencies range between 4.5 and 8 cps. The increased number of responses at frequencies up to 8 cps are those related to the thoraco-abdominal system. This demonstrates further that under these acceleration conditions the sensations experienced are dependent upon the vibration frequency involved.

Each ride provided a history of sensations that were rated for intensity on a 0–4 scale. Table 5-2 lists the results of these subjective estimations. The regional symptomatology given in the left column contains

TABLE 5-2 *Average Subjective Ratings on 0 to 4 Scale of Sensations Experienced during Whole-body Vibration*

Body-region symptomatology	Average subjective rating per frequency										
	1 cps	2 cps	2.5 cps	3 cps	3.5 cps	4 cps	4.5 cps	5 cps	5.5 cps	6 cps	7 cps
Head-neck area:											
Head sensations—vibration or "tight sensation" of facial skin	0	0.1	1.0	1.3	1.1	1.7	1.3	2.7	3.6	3.4	3.7
Pharynx—pharyngeal tug or "lump in throat" (occurs only above 10 cps)											
Jaw—sensation of vibration (occurs only above 10 cps)											
Speech—lower frequencies secondarily affected because of reactions of thorax and abdomen; high frequencies affected because of superimposed transmitted vibrations to laryngeal tissues and possibly main-stem bronchi	0	0.4	0.3	0.7	1.0	2.2	1.5	2.4	2.8	3.6	3.7
Thorax:											
Respiration—decreased ability to perform physiological respiratory movements of the thoracic cage because of superimposed forces from oscillating platform	0.2	0.7	1.0	1.2	1.7	1.9	2.7	3.1	2.9	3.7	3.9
Dyspnea—air hunger	0	0.5	0.6	0.1	0.6	1.2	1.2	2.7	3.5	3.4	3.8
Valsalva—partial or complete closure of glottis, resulting in increased intrathoracic and intra-abdominal pressure	0	0.4	0.7	0.5	0.7	1.6	1.4	2.8	3.3	3.3	3.8
Substernal and/or precordial pain—dull to severe pain of the precordium, occasionally radiating to the sternum but with no other radiations; pain subsides immediately upon cessation of vibration	0	0	0	0	0	0	0	1.8	2.0	3.2	3.4

TABLE 5-2 *Average Subjective Ratings on 0 to 4 Scale of Sensations Experienced during Whole-body Vibration (Continued)*

Body-region symptomatology	Average subjective rating per frequency										
	1 cps	2 cps	2.5 cps	3 cps	3.5 cps	4 cps	4.5 cps	5 cps	5.5 cps	6 cps	7 cps
Abdomen:											
Voluntary abdominal musculature contraction—degree of contraction or "bearing down"	0	0	0	0.3	0.8	1.3	2.9	3.2	2.8	3.3	3.5
Abdominal pain—usually periumbilical with tendency to radiate to right lower quadrant	0	0	0	0	0	0	0.6	2.0	1.3	2.5	1.9
Skeletal musculature:											
Skeletal musculature—sensation of "muscle tightness" or possibly increased muscle tone primarily of the lower extremities, dorsum, and neck (occurs only above 10 cps)											
Voluntary muscle contraction of extremities—muscular contraction in an effort to counteract severe movements of oscillating platform	0	0	1.0	0.3	1.8	1.8	1.6	3.0	3.4	3.6	4.0
Lumbosacral pain—dull to severe pain at midline with bilateral radiation; pain subsides immediately upon cessation of vibration	0	0	0	0	0	0	0.4	0.8	0	0.4	0.4
Pelvic-perineal complex:											
Micturate (urge)—mechanical stimulation of bladder neck and proximal portion of urethra (occurs only above 10 cps)											
Defecate (urge)—mechanical stimulation of distal portion of sigmoid colon and rectum (occurs only above 10 cps)											
General discomfort	0	0	1.0	0.5	0.6	1.7	1.1	2.8	3.4	3.7	3.9
Total (overall)	0.2	2.1	5.6	4.9	8.3	13.4	14.7	27.3	29.0	34.1	36.0
Total (minus pain)	0.2	2.1	5.6	4.9	8.3	13.4	13.7	22.7	25.7	28.0	30.3

not only those symptoms reported in response to 1 to 7 cps but also those for 8 to 20 cps. Ratings are not available for the latter frequencies.

THE RELATION BETWEEN SENSATIONS AND PHYSICAL RESPONSE

Obtaining subjective response is one method that may be used to define certain mechanical reactions of the body to extrinsically applied vibrations. The use of this technique is best illustrated by comparing the experimental results with the thoracoabdominal system and the subjective responses. Resonance of the thoracoabdominal system was mechanically measured and found to be at approximately 3 cps when the body is in supine position with relaxed abdominal muscles [5]; however, intensive abdominal pain was experienced between 4.5 and 10 cps. Subjective responses suggest that resonances in these regions change as body position is changed and that subsystems, not discernible by external mechanical measurements, within the thorax may be excited by vibration.

The etiology of symptoms is determined by the physiological alterations brought about by the mechanical stimulus to whole-body vibrations. In turn, the resultant mechanical stimulus is dependent upon the mechanical properties of organ-tissue complexes. Therefore, subjective response under these conditions is considered to be a physiological manifestation of the direct or indirect mechanical excitation of various organ-tissue complexes and the resulting stimulation of associated sensory receptors. The sensory receptors make up a profuse complex that maintains general somesthetic sensibility. It is assumed that vibration per se affects only those receptors susceptible to displacement. These include the mechanoreceptors of touch, pressure, and proprioception; pain which may result from tissue displacement is also included. Since both mechanoreceptors and pain receptors are involved in producing sensations, they are grouped together and referred to as "kineceptors." These receptors represent a continuum of sensations from displacement to pain and constitute an excellent informative and protective mechanism for the organism.

The various sensations encountered during the tolerance studies were grouped and are shown in Figure 5-17, which illustrates the symptoms or sensations experienced at or near tolerance levels within the frequency range of 1 to 20 cps. Up to approximately 10 cps, those sensations referable to the thoracoabdominal system are experienced; those sensations referable to skeletal-musculature tonus, head, and perineum are experienced between approximately 10 and 20 cps. Thus, those structures having lower resonances (more freedom of movement and greater mass) would be most affected at the lower frequencies, and those structures that have higher resonances (little freedom of movement and less mass) would be most affected at the higher frequencies. It is evident, then, that the

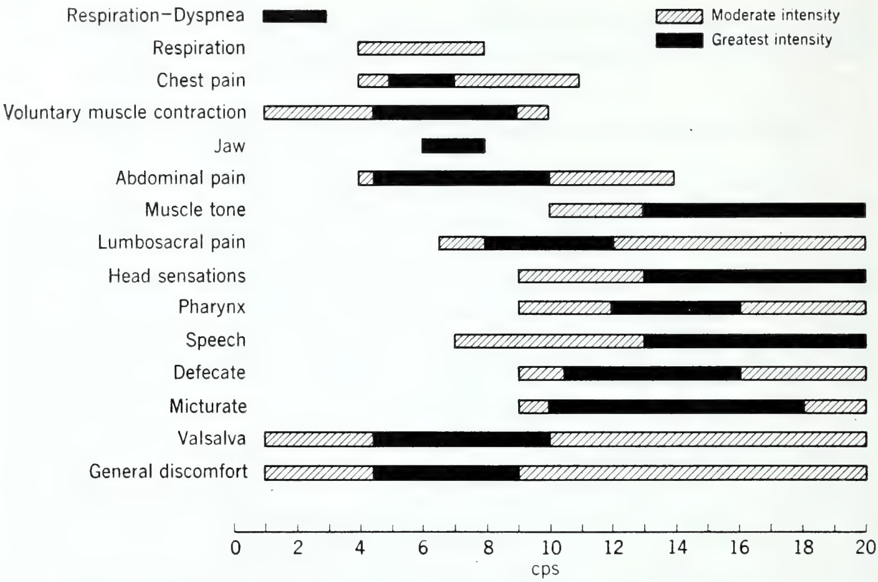


FIG. 5-17 *Range of perceptible sensations. (After E. B. Magid and R. R. Coermann, The Reaction of the Human Body to Extreme Vibrations, Proc. Inst. Environmental Sci., p. 135, 1960. By permission of the publishers.)*

phenomenon of resonance plays a dominant role in determining subjective tolerance and tissue-organ tolerance to whole-body vibration.

Head and Neck

The mandible and its surrounding soft tissues apparently resonate in the range of 6 to 8 cps; the head sensations were first experienced at approximately 9 cps. The subject's first impression of head sensations was a feeling of vibration of the facial skin, particularly about the cheeks and eyelids. This sensation appeared to increase in violence up to approximately 13 cps, after which it subsided considerably and was replaced by a sensation of "tightness." The entire head was enveloped by this sensation, and increased effort was required for moving the head, eyelids, and jaw. This tightness may have been caused not only by skin movements but also by increased muscle tone of the facial musculature. Most subjects complained of a low-grade full-frontal headache (which usually lasted for a few minutes following the run) during exposure to the higher frequencies. The etiology of this head pain has not yet been determined. It is of interest and perhaps necessary to know what forces are transmitted to the brain, meninges, vasculature, and cerebral-spinal fluid and how these structures and fluid react to the transmitted forces.

The sensation of the pharynx was described as a "tug" or "lump in the throat," which is similar to the description given for "globus hystericus." It began at 9 cps and had its greatest intensity between 12 and 16 cps. Apparently the trachea and perhaps the main-stem bronchi resonate at these frequencies, and the sensations may have been the result of mechanical stimulation by the trachea of the sensory receptors of the supporting pharyngeal tissues. The sensation of choking or stimulation of the cough reflex never occurred during the experiment.

Speech appeared to be most affected between 16 and 20 cps. At the lower frequencies, speech was secondarily affected by reactions of the thorax and abdomen. At the higher frequencies, however, the superimposed transmitted vibrations to the trachea, particularly those acting at the laryngeal area, had a greater detrimental effect upon phonation.

Thorax

The thorax contains the pulmonary and cardiac systems vital to body integrity. The time period in which these systems may be adversely affected is extremely critical. The systems are integrated in such a fashion that effects on one system have a direct and immediate effect on the other. Because the thorax contains highly mobile structures with a large gas-solid interface and its structures have viscoelastic characteristics, the greatest displacements occur at the lower frequencies. Between 1 and 3 cps, the forces due to the large excursions necessary for approaching tolerance become superimposed upon and considerably impair physiological respiratory movements. The subjects experienced breathing difficulty and the sensation of air hunger. Changes in pulmonary circulation probably contributed to dyspnea. Between 4 and 8 cps the respiratory movements were less modified by the superimposed movements of the vibrations, and dyspnea was not experienced.

Chest pain was experienced between 4 and 11 cps. At 4 to 9 cps the pain occurred in the left chest and had a distribution similar to that of coronary heart disease; the pain was dull and aching in nature and increased in intensity as the ride continued. More than half the subjects complained of dull, low-grade, intermittent precordial pain lasting up to 7 days. There were no radiations to the left shoulder or down the left arm. This pain may well have originated from mechanical stimulation of the heart on the diaphragmatic pericardium and on the parietal pericardium about the base of the heart. Also it is probable that stretching or deformation of the major vessels, particularly the aorta, and their supporting structures played a role in producing pain. At 9 to 11 cps the pain was experienced in the substernal area, radiating out bilaterally, and was increased by respiratory movements. It is believed that this

pain originated at the costosternal articulations from displacements of the thoracic cage; displacement of the anterior attachments of the diaphragm could also produce this pain.

Abdomen

Abdominal pain occurred between 4 and 14 cps and was usually distributed about the umbilicus and radiated down to the right lower quadrant. The pain is believed to have had its origin in the stretching and deformation of the terminal ileum, cecum, and hepatic flexure of the large intestine and its supporting mesenteries. Above 8 cps the pain has a definite tendency to be located between the umbilicus and the pubic symphysis.

Pelvic-Perineal Organs

At 9 to 20 cps there occurred a definite sensation of the urge to defecate and urinate. Prior to each run, the subjects were instructed to empty their bladders and to defecate if possible so that at the beginning of each run the pelvic-perineal organs would be empty. At 10.5 to 16 cps the urge to micturate was extremely intense, and the subjects could hold back only with extreme effort. The urge to defecate followed the same pattern. Stretching and deformation of the walls of these potentially hollow organs, perhaps because of resonance, stimulated the sensory receptors within these tissues.

Skeletal Musculature

Voluntary muscular contraction occurred primarily between 1 and 10 cps. At these frequencies the subjects tended to bounce in the seat no matter how tightly the lap belt and shoulder straps were fastened. In addition, the abdomen, thorax, and extremities were alternately forcibly displaced, and, to compensate for this effect, the subject had to hold tightly the handles of the seat; simultaneously the lower extremities were forcibly pushed against the vibrating platform. Voluntary muscular contraction reached its maximum intensity between 4.5 and 9 cps.

Above 10 cps the movement of the platform smoothed out because of the increased frequency and decreased amplitude. The subject then experienced an involuntary "tightening" sensation of the lower extremities, back, neck, and head, which became extremely intense beyond 13 cps. It is suggested that at these accelerations the transmitted forces and perhaps the resonances of the skeletal musculature increased the

mechanical stimulation of the myotatic receptors and thereby increased muscle tonus. During whole-body vibrations at tolerance levels, there was a marked increase in stimulation of mechanical receptors of the sensorium. At present the effect of massive afferent input to the spinal cord, in terms of central excitatory and inhibitory states, can only be surmised. The patellar reflex has been observed to decrease during high-frequency vibration [1]. Therefore, it is probable that an increase in myotatic stimulation and spinal-cord response play a role in this phenomenon. It is apparent that performance is affected by alternating displacement of body parts; the change of skeletal-muscular tonus must also be considered.

Lumbosacral Area

Lumbosacral pain was experienced by at least four subjects in the range of 6.5 to 20 cps. Between 8 and 12 cps eight subjects experienced moderate to severe pain in the low midback region. The pain was dull in nature, slowly increasing in intensity as the ride continued. The subjects reacted to this by muscle guarding of the area. The pain subsided immediately after the run; no sequela was reported.

Valsalva Maneuver

A partial or complete valsalva maneuver was performed with reflex-like characteristics for all frequencies at tolerance levels. This maneuver consists of a forced expiration against a closed or partially closed glottis; it is accomplished by forcibly contracting the thoracic and abdominal musculature and the diaphragm. If the glottis is closed, gas present in the lungs cannot escape and there results an increase in positive pressure transmitted to the thoracic and abdominal compartments. This has the effect of stiffening the walls of the compartments, particularly of the abdomen, and compressing the organs together. Performing this maneuver decreases displacement of the viscera because of the increase in stiffness and damping, which increases protection against alternating forces. Earlier it was shown that impedance decreased with the application of a semirigid envelope about the abdomen.

The maneuver was always performed when pain or respiratory difficulties were experienced. The greatest effort in performing the maneuver occurred between 4.5 and 10 cps. This coincides with the frequencies having the greatest effect upon the abdomen and thorax. The valsalva maneuver and voluntary muscular contraction are the two techniques whereby the subject attempted to protect himself from extreme displacements.

The valsalva maneuver has several cardiac and pulmonary effects. The slightly negative intrathoracic pressure normally present during inspiration may increase by several millimeters of mercury and result in decreased venous return to the right atrium. This might be compensated by the effect of venous massage caused by alternating forces and increased muscular activity. It has been shown that exercise, pain, and apprehension, which are experienced during vibration, can increase venomotor constriction and lead to increased central and peripheral venous pressures [1]. For the time period during and following its performance, the valsalva maneuver is the stimulus causing the greatest venomotor activity. It would be interesting to determine the effect of vibration on this mechanism. The subject could be adversely affected if cardiac output were to be decreased to a critical level for any length of time, particularly since there is an increased demand for blood because of the increased muscular activity during these oscillations.

Partial valsalva consists of expiration against a partially closed glottis. The net result is similar to the condition found in positive-pressure breathing. Increased intrapulmonic pressure necessitates increased work of the right heart, which results in pulmonary hypertension. Therefore, intermittent long-term vibrations of the magnitude studied might have deleterious physiological effects. It would be of interest to determine the effect of these severe alterations on hemodynamics, with particular attention to respiratory gases, and the effects of these conditions on vagal activity.

General Discomfort

General discomfort refers to an overall subjective response given for each frequency. Pain, respiratory difficulties, muscular effort and urgencies, although of varying intensities, may be experienced simultaneously for more than one frequency. Therefore, the subject's general impression of each ride was ascertained to determine the willingness to be vibrated at any given frequency. As would be expected, somatotype has a definite effect on symptoms. For example, the wide-chested individual having a horizontally rotated heart experiences chest pain earlier and the pain is of greater intensity than the pain experienced by a narrow-chested individual having a vertically rotated heart. It should be mentioned that at no time was dizziness or vertigo experienced.

CENTRAL VS. PERIPHERAL FREQUENCY RESPONSE

A comparison of mechanical and subjective response is presented in Figure 5-18. The greatest total subjective response and the least tolerance

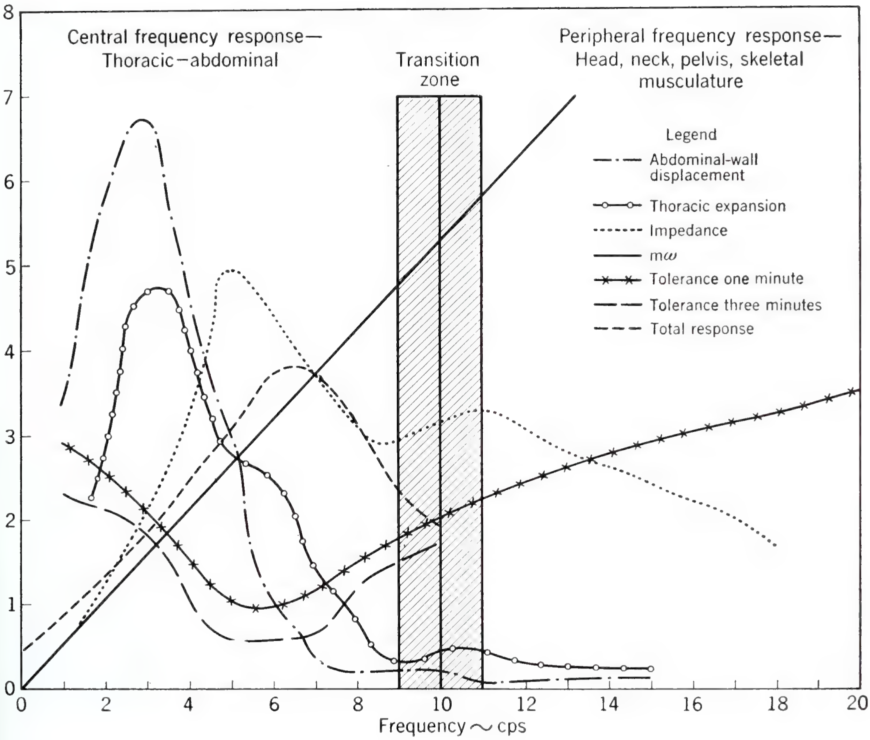


FIG. 5-18 *Comparison of mechanical and subjective response.*

to vibrations occur within the frequency range of 1 to 10 cps. In this range, the greatest mechanical response for the thoracic and abdominal compartments correlates with the greatest physiological response of these areas and thus with the greatest subjective responses. The frequencies of greatest thoracoabdominal response to vibration may therefore be referred to as the central frequency range.

The symptoms referable to the head-neck, pelvic-perineal, and skeletal-muscular systems occur primarily in the frequency range of 10 to 20 cps (within the scope of the present studies). Rigid skeletal structures are inherent to each of these systems. Resonances of these structures would be expected to occur within the higher frequency range. Indeed, the symptom spectrum demonstrates that these structures are mechanically excited, perhaps because of resonance above 10 cps. The frequencies having the greatest mechanical and therefore the greatest physiological effect upon these body regions are referred to as the peripheral frequency range.

The concept of a central and a peripheral frequency range provides a means of categorizing the occurring phenomena and may aid in future

investigations of extrinsic forces. For example, mechanical impedance is of greatest magnitude within the central frequency range, which is the area of greatest subjective response (most irritating and painful) and minimal tolerance. This suggests that the area of greatest physiological stress to the individual is within the range of greatest mechanical impedance. If this concept is valid, perhaps it can be applied to human and animal experimentation alike; the frequencies of greatest stress would be determined by the mechanical impedance and would be used to define the frequency range that should be studied. This technique may aid in comparisons between human and animal studies.

PHYSIOLOGICAL RESPONSE

Measurements of physiological alterations under dynamic conditions are not only extremely difficult but potentially hazardous to human subjects. The traditional clinical and laboratory techniques that have been developed to function under ideal static conditions are almost useless, since the instrumentation is usually adversely affected by dynamic forces. Therefore, it is necessary to develop specialized techniques and instrumentation for the observation of specific physiological phenomena. The following are general descriptions of studies that have been performed on physiological response.

Cardiovascular Response to Whole-body Vibration

By virtue of its location, anatomical configuration, and function, the cardiovascular system, which is vital to moment-to-moment body integrity, is extremely vulnerable to certain extrinsic mechanical forces. As has been observed in experiments, chest pain, dyspnea, increased muscular activity, and valsalva maneuver indicate physiological alterations of this system.

Blood-pressure Measurements

Blood pressure was measured directly by radial artery puncture within the frequency range of 2 to 20 cps at tolerance accelerations for a 1-min period. It was observed that from 2 to 5 cps the blood pressure was almost double the control levels, whereas from 6 to 20 cps blood pressure generally decreased by as much as a third of the control levels. Pulse pressure generally showed little change.

At the lower frequencies there occurred severe alternating dis-

placements of the thorax and abdomen, partial valsalva was performed, and considerable voluntary muscle contraction was exerted. It is probable that at these frequencies the multiple mechanical body reactions acted as an efficient venous "pump" that vigorously massaged the great veins and thereby increased venous return. This effect was reflected in the significant rise of the diastolic pressures. Above 5 cps, pain of the thoracoabdominal compartments, complete valsalva, general discomfort, and the sensation of increased muscle tonus were experienced. The result was a significant drop in systolic and diastolic pressures. Frequencies below 5 cps produce radically different effects.

Electrocardiographic Observations

The heart was monitored directly by means of electrocardiographic tracings. The tracings were taken before, during, and after each run. In only one case were abnormal tracings observed. After a 1-min run at 8 cps the subject experienced diaphoresis and lightheadedness associated with momentary syncope. This effect was associated with inversion of the P wave and tachycardia. The syncope was transitory, and the P wave reverted, with no sequela after approximately 2 min. It was later observed that the subject's previous electrocardiograms showed an occasional inverted P wave following a Masters two-step test.

Among 15 subjects tested, heart rate was observed to increase 10 to 15 beats per minute during all runs, regardless of frequency. Although relative hypertension occurred at the lower frequencies and relative hypotension was caused at the higher frequencies, heart rate generally differed little among frequencies. This suggests that different compensatory mechanisms become prominent under these conditions and result in significant mechanical changes but insignificant electrical changes with the techniques used. This study is an excellent illustration of the limited value of the electrocardiogram in monitoring cardiac dynamics in subjects undergoing hazardous mechanical manipulations.

Red- and white-blood-cell counts and white-blood-cell differentials were observed during vibration. No significant changes were found immediately after or up to several hours following vibration. Sedimentation rate also showed no significant changes.

Respiration during Vibration

At a constant acceleration of approximately 1.3 g for a 1-min period with a frequency range of 3 to 15 cps, respiratory mechanics and oxygen uptake were studied in seven subjects. Tidal volumes and therefore minute volumes nearly doubled for all frequencies except at 5 to 7 cps,

at which the values were at least four times greater than the original volumes. Vital capacity generally showed slight decreases between 5 and 7 cps. Maximum breathing capacity decreased by at least 40 per cent at 7 cps. Oxygen uptake generally doubled, but at 5 and 7 cps it nearly trebled.

It has already been shown that the thoracoabdominal compartments resonate at approximately 3 cps for subjects in a supine position with relaxed abdominal musculature. However, it was found that the greatest respiratory changes usually occurred between 5 and 7 cps, thereby showing that factors other than the normal resonance predominate. Chest and abdominal pain, vigorous voluntary muscular contractions, particularly of the extremities, and valsalva maneuver were strenuously performed between 5 and 7 cps. Apparently, one important effect of these dynamics is the change in the natural frequency of the thoracoabdominal system. Study of the effects of these severe reactions on cardiopulmonary dynamics is required for the full understanding of the physiological alterations that occur.

Endocrinological Response

At the end of a series of runs the subjects experienced facial flush, diaphoresis, and euphoria. All subjects felt extremely relaxed. Within 2 to 4 hr after the run, most subjects experienced weariness and depression, which lasted for several hours. These findings indicate alterations in hormonal secretions during and following vibration.

Space flight involves mechanical forces during entry, reentry, and transient accelerations separated by long periods of weightlessness. Therefore, consideration should be given to possible endocrine and metabolic changes secondary to mechanical stimuli and the effects of these on diurnal cycling in man. In order to maintain space-crew effectiveness, particular attention should be given to the work-rest cycle [17].

Performance during Whole-body Vibration

Mechanical reaction of the body to vibration interferes directly with physical activity. Physiological alterations, including subjective response secondary to the mechanical response, may also interfere with physical activity. It is, therefore, of great importance to anticipate performance decrements that may be caused by the influence of severe noxious alternating forces.

One of the most important modalities associated with performance is that of vision. Sharp decrements of visual acuity were observed in

subjects within the frequency range of 40 to 100 cps at comparatively low g levels [1]. The dependence of acuity upon frequency was interpreted as being due to resonances of the eyeballs. Visual acuity was further studied within the frequency range of 1 to 7 cps at a ½-g vector. At 3 cps acuity was decreased by as much as 30 per cent, and at 6 and 7 cps the decrement was as high as 35 per cent. Further studies at 3-min tolerance levels indicated constantly decreasing acuity at 10 to 20 cps. The factors having the greatest adverse effect on visual acuity at certain frequencies were found to be the mechanical interference due to the mechanical characteristics of the body and the secondary subjective responses experienced at tolerance levels.

Equilibrium during Vibration

An "equilibrium chair" was designed for measuring the ability of a subject to maintain the horizontal position in pitch and roll while he underwent vibrations at 1 to 20 cps at 1-min tolerance levels. The chair was installed on a shake table and designed to produce random, simultaneous changes in pitch and roll that could be compensated by a control stick installed in the chair. It was the task of the operator to compensate for all motions and to maintain what was thought to be the horizontal position [2]. Under these conditions, it was found that some subjects completely lost control of equilibrium during vibrations, whereas others appeared to be only slightly affected by the vibrations. One subject developed severe, sharp pain in his left upper quadrant during the first few runs and could not continue the study. The only abnormality in his clinical history was an acute episode of malaria 10 years previously, with an associated splenitis.

Further work showed that the ability to maintain the horizontal position during vibration was significantly affected between 4 and 10 cps, with the greatest decrement occurring between 5 and 8 cps. This decrement between 5 and 8 cps persisted after vibration, demonstrating significant lasting physiological alterations following exposure to these noxious forces.

CONCLUSIONS

High-performance jet-propulsion systems in aerodynamic and space vehicles challenge man with a new and highly complex, potentially lethal, poorly understood environment. Vibration is one of the mechanical forces to be encountered. At present the major problems of vibration are confined to low-frequency high-amplitude oscillations involved in rocket

burnout, transient accelerations, and reentry phases of space flight. Because of the mechanical characteristics of the body, the frequency range that requires particular attention is that from 0.5 to 20 cps.

The studies discussed in this chapter demonstrate significant mechanical and therefore biological phenomena that last for relatively short periods of time. It has been shown that human beings are adversely affected in the frequency range of 1 to 20 cps and are particularly vulnerable in the range of 1 to 10 cps. Subjective response (including severe pain) and cardiovascular, respiratory, skeletal-muscular, and performance alterations are among the various effects of these extrinsically applied environmental forces. If these noxious forces are to be encountered, the acute and chronic effects on the health of the passenger must be anticipated. These observations are the result of investigations of carefully controlled short-time steady-state sinusoidal vertical vibrations with a specific seating and restraint configuration. It is necessary to extend this work to the study of long-term states, intermittent buffeting, and single repetitive impacts. Also needed is the investigation of combined multidirectional forces with varying seating and restraint systems.

BIBLIOGRAPHY

1. Coermann, R. R.: The Effect of Vibration and Noise on the Human Organism, *U.S. Dept. Commerce Publ.* PB-24679-T, 1938. (Translated from the German.)

2. ——— and E. B. Magid: The Influence of Vibration on Holding the Horizontal Position Utilizing the Equilibrium Chair: Project Mercury Candidate Evaluation Program, *USAF WADC Tech. Rept.* 59–505, December, 1959.

3. ——— et al.: The Passive Dynamic Mechanical Properties of the Human Thorax-Abdomen System and of the Whole Body System, *Aerospace Med.*, vol. 31, pp. 443–455, June, 1960.

4. Dieckmann, D.: Einfluss Vertikaler Mechanischer Schwingungen auf den Menschen, *Intern. Z. angew. Physiol. einschl. Arbeitsphysiol.*, vol. 16, p. 519, 1957.

5. DuBois, A. B., et al.: Oscillation Mechanics of Lungs and Chest in Man, *J. Appl. Physiol.*, vol. 8, pp. 587–594, 1956.

6. Fishbein, W. I., and L. C. Salter: The Relationship between Truck and Tractor Driving and Disorders of the Spine and Supporting Structures, *Ind. Med.*, vol. 19, pp. 444–445, 1950.

7. Fowler, R. C.: Damage to Animals Due to Vibration, Supplement to *USN Research Lab. Bull.* 22, Shock and Vibration, 1955.

8. Gierke, H. E. von: Transmission of Vibratory Energy through Human Body Tissue, *Proc. First Natl. Biophys. Conf., 1957*, pp. 647–668, Yale University Press, New Haven, Conn., 1959.

9. Goldman, D. E.: A Review of Subjective Responses to Vibratory Mo-

tion of the Human Body in the Frequency Range 1 to 70 cps, Project NM 004 001, Naval Medical Research Institute, Bethesda, Md., 1948.

10. Guillemin, V., Jr., and P. Wechsberg: Physiological Effects of Long Term Repetitive Exposure to Mechanical Vibration, *J. Aviation Med.*, vol. 24, pp. 208–221, 1953.

11. Hunt, J. H.: Raynaud's Phenomenon in Workmen Using Vibrating Instruments, *Proc. Roy. Soc. Med.*, vol. 30, pp. 170–171, 1936.

12. Lenke, J. P.: Health Hazards from the Use of the Air Hammer in Cutting Indiana Limestone, *U.S. Bur. Labor Statist. Bull.* 236, ser. 19, July, 1918.

13. Loriga, G.: Pneumatic Tools, in "Encyclopedia of Hygiene, Pathology and Social Welfare," vol. 2, Occupation and Health, pp. 270–274, International Labor Office, Geneva, Switzerland, 1934.

14. Magid, E. B., R. R. Coermann, and R. Lowry: Comparison of Symptomatic Response to Mechanical Properties of the Body during Sinusoidal Vibration, unpublished paper.

15. ———, ———, and G. H. Ziegenruecker: Human Tolerance to Whole Body Sinusoidal Vibration, *Aerospace Med.*, vol. 31, pp. 915–924, November, 1960.

16. Paulson, E. C.: Tractor Drivers' Complaints, *Minn. Med. J.*, vol. 32, p. 386, 1949.

17. Ray, T. R., and E. O. Martin, Jr.: Human Performance as a Function of the Work-Rest Cycle, *Natl. Acad. Sci.–Natl. Research Council Publ.* 882, 1961.

18. Riopelle, A. J., Marion Hines, and M. Lawerance: The Effects of Intense Vibration, *USA Med. Research Lab. Rept.* 358, 1958.

19. Sueda, M.: The Effect of Vibrations on the Estrus Cycle of White Rats, *Mitt. Med. Akad. Kioto*, vol. 21, pp. 1066–1082, 1937.

20. Ziegenruecker, G. H., and E. B. Magid: Short Time Human Tolerance to Sinusoidal Vibrations, *USAF WADD Tech. Rept.* 59-391, July, 1959.

Chapter 6

BIOLOGIC DOSIMETRY

*Gerrit L. Hekhuis**

THE ADEQUATE EVALUATION OF ANY OBSERVED EFFECT WITH REFERENCE to a precedent cause depends critically upon the accuracy of measurement of both the cause and the related effect. The evaluation of the effect of radiation on a biologic system certainly has not been limited by a lack of biologic effects or end points that have been observed, recorded, and reported. Rather, the relation between observed changes and prior incident radiation has been obscured by a lack of information concerning the exact magnitude, or dose, of the radiation and its location.

RADIATION CHARACTERISTICS

When the effects of high-energy ionizing radiation on matter are considered, two essentials to understanding include a terminology commonly accepted and a familiarity with the various ways in which these radiations interact with the material exposed.

As yet, universal agreement has not been reached on the nomenclature for many of the quantities used in radiation measurement, or dosimetry. Perhaps the best guide for consistent usage is that provided by the International Commission on Radiological Units and Measurements (ICRU) [9]. An appreciation of the problems of dosimetry requires an acquaintance with a few elementary concepts of atomic physics, which are covered very briefly in this discussion. For an understanding or application of the principles, the reader is referred to the many excellent texts in this field [1 to 3, 6].

The Radiation Beam

The radiations of biologic concern may arise from many types of sources. Nuclear transformations, particle accelerators, and cosmic space

* DCS/Operations, Aerospace Medical Division, Air Force Systems Command, Brooks Air Force Base, Tex.

This chapter represents the views of the author and does not necessarily represent the views or sanction of the U.S. Air Force.

serve as sources for these radiative emanations. Included in this radiation are subatomic and atomic particles and electromagnetic radiations in the form of X and gamma rays. Any given beam of these radiations has certain physical characteristics. Among these are the energies and the intensities of the photons and particles.

Energies of X rays and gamma rays, which are electromagnetic radiations differing mainly in their origin, are expressed in terms of the energy of each of the photons, E_p, which is then defined as

$$E_p = h\nu = \frac{hc}{\lambda} \tag{6-1}$$

where h = Planck's constant
c = velocity of light
ν = frequency
λ = wavelength

The energies of gamma rays from nuclear changes typically range from a few thousand electron volts (kev) to several million electron volts (Mev). X rays, produced from interactions involving electrons, range in energy from a few electron volts (ev) to very high levels (from high-energy particle accelerators) in the billion-electron-volt (Bev) region.

Intensities of X and gamma rays are expressed as energy per second in the form of photons passing a plane perpendicular to the direction of motion. Since this plane area of reference transects a solid angle cone, some positional variation in intensity is to be expected. Thus, the intensity of electromagnetic radiation can be thought of as the energy per second as photons pass from any direction through a sphere of unit cross-section area.

Beams of charged-particle radiations can also be expressed in terms of intensities, by using the *number* of particles crossing a unit area rather than the energy. Flux is analogous to spherical density.

Neutron beams can be designated similarly; they are commonly expressed in the number of neutrons of specified velocity or energy per square centimeter per second.

Interactions

The manner in which radiation interacts with matter through which it passes is essentially mediated by the transfer of some of the energy from the radiation (loss of energy) to the material radiated (excitation, ionization). The interactions vary somewhat with the different radiations, but when the main concern is with the result of the energy transfer, or absorption, the similarities of action are determined mainly by charge and mass.

Since all radiation measurements depend on this interaction of radiation with matter, it is necessary to typify a few of these interactions.

An electromagnetic radiation (X- or gamma-ray) tends to lose much or all of its energy in a single interaction. These interactions are of two principal kinds: the photoelectric effect, which predominates at intermediate energies (0.05 to 15 Mev in aluminum, 0.5 to 5 Mev in lead), and pair production, which predominates at high energies [11].

In these kinds of interactions, the net results are transfer of energy resulting in excitation, activation, or ionization of the atom and absorption or deflection of the incident radiation as a result of its interaction. With each successive interaction, then, the intensity and flux characteristics of the beam are being changed.

This change of the beam as it passes through and reacts with matter is of considerable concern to the biophysicist charged with determining the characteristics of the radiation incident upon a specific, or perhaps unknown, volume within a biologic system.

Attenuation of an electromagnetic radiation beam passing through matter follows the classic exponential relation, but strictly only if the beam is extremely narrow. This relation of the intensity of a monoenergetic narrow beam before and after passing through an absorber of thickness x is expressed by

$$I = I_0 e^{-\mu x} \tag{6-2}$$

where I_0 = intensity with no absorber
μ = linear absorption coefficient

Often it is useful to express the absorption of radiation in terms of the thickness of absorber required to reduce the incident intensity by 50 per cent. This thickness of material, the half-value layer (HVL) in a narrow beam undergoing exponential absorption, is

$$\text{HVL} = \ln\frac{2}{\mu} = \frac{0.693}{\mu} \tag{6-3}$$

The above discussion applies only to monoenergetic narrow beams of photons. Where many energies are mixed, the absorption coefficients must be related and the net absorption coefficient is determined by the spectrum of the energies. As the absorber is increased in thickness, the most easily absorbed components of the beam are removed, leaving a "harder" beam with a smaller effective absorption coefficient. Absorption is also modified by the density of the material; this requires use of the mass absorption coefficient μ/p, which is expressed in grams per square centimeter, and the intensity relation becomes

$$I = I_0 e^{-(\mu/p)px} \tag{6-4}$$

Contrary to the tendency for electromagnetic radiation to lose energy in a very few interactions, charged particles lose their energy in a large number of small energy transfers, chiefly through Coulomb interaction with atomic electrons.

Depending on the amount of energy transfer at each atomic encounter, the process is designated as excitation or ionization. Together, this energy loss is termed collision. As a result of this interaction, some of the ejected electrons are themselves capable of further ionization, and these electrons are called delta rays.

Some interactions of charged particles with matter involve the Coulomb fields of the atoms, and particularly of the nuclei, resulting in a change of direction of the incident particle. If these involve no kinetic energy change to another form, this interaction is called elastic scattering. For electrons, scattering through a large angle involves an appreciable acceleration, and this in turn may result in the emission of a photon of electromagnetic radiation, called *Bremsstrahlung* [11].

Since in these reactions the ionization produced involves electrons that are bound to atoms, the energy transfers are confined to finite distances from the track of the moving primary particle. This results in a roughly cylindrical volume of ionization coaxial with the particle track. The radius of this cylinder depends on the range of energy transfer, which is modified by the atomic number of the absorber and the speed of the particle.

The density of ionization varies both from the central axis out to the surface of the cylinder and along the linear path of the particle, in terms of frequency of interacting events. This ionization density along the linear path is important in volume-dose estimation and is related to the linear energy transfer (LET).

Linear energy transfer is defined as "the linear-rate loss of energy (locally absorbed) by an ionizing particle traversing a material medium."[1] LET may be expressed conveniently in kilo electron volts per micron. Thus, the employment of the LET concept emphasizes the transfer of energy to the medium, rather than the loss of energy by the particle, and thus is useful to the biologist, who finds that the biological effect of radiation depends on the spatial distribution of the energy transfer as well as the total energy transferred per gram. Since the energy transfer involves a range from some maximum value down to zero, the LET is also variable through a range. Thus, in calculations, either an average value of sorts or the complete distribution derivation must be used.

Once the linear energy transfer in energy per unit path is known, it is possible to calculate the number of ionizations expected if the energy per ionization is known. In a study of the energy loss per ion pair

[1] ICRU report [9], pp. 1–2.

produced, Gray [5] reported values from electrons, protons, and alpha particles in air, hydrogen, and helium of 31.0 to 36.0 ev per ion pair. The ICRU [9] recommends that the energy per ion pair be taken as 34 ev.

Associated closely with the rate of energy loss along its path is the resulting maximum range of the charged particle. Obviously the particle will cease motion when the available energy has been expended. These finite ranges, dependent on energy, are calculable by pertinent formulas, which take into consideration the kinetic energy, charge, velocity, density of absorber atoms, atomic number of the absorber, etc. For alpha particles, in a material of atomic weight A, the range,

$$R_A \text{ (mg/cm}^2\text{)} = 0.56R \text{ (cm)} A^{1/3}$$

where R = range of alpha in air at standard temperature and pressure. For electrons with energy E between 0.8 and 3.0 Mev, the range

$$R_e \text{ (mg/cm}^2\text{)} = 0.542E - 0.133$$

If the electron energy ranges from 0.15 to 0.8,

$$R_e \text{ (mg/cm}^2\text{)} = 0.407E^{1.38}$$

These empirical formulas are quoted by Price [8] from data of Glendenin [4].

DOSIMETRY

With only the previous hints at the complexity of the incident radiation beam, it is now necessary to consider the problem of measuring the radiation of concern at some point of interest. In surveying the available measurement devices it soon becomes evident that there is no single solution.

A few of the factors complicating the problem are energy dependence of the detecting material, equivalence of ionization in the material to ionization in air or tissue, electronic reaction or discrimination, geometric relation of the source and measuring device, determination of intensity and energy spectrum of the beam at some depth within an absorber, estimation or calculation of the effective density of the material, and the related linear energy transfer with its variations.

Only a brief summary of radiation-measurement devices can be afforded, and the student is referred to detailed text and current literature discussions of individual methods.

Roughly, the different types of measuring devices can be categorized by the type of interaction of the radiation with the detector element. Several types depend on a collecting volume within which ionization

is produced by the passage of the charged particles. Instruments of this type include Geiger-Müller counters, ionization chambers, proportional counters, crystal counters, and cloud chambers. These instruments can be used under certain conditions for X and gamma rays and uncharged particles such as neutrons by employing means to effect ionization by secondary processes.

In other methods, utilization is made of properties of excitation and, sometimes, molecular dissociation. These properties, coupled with ionization, give rise to the usefulness of scintillation counters. The use of photographic-emulsion latent images is also an application. Molecular dissociation is the basis for many of the liquid-chemical detection and dosimeter techniques.

Other methods, involving radiative interactions, are somewhat more specific and are undergoing a most rapid investigative and development emphasis.

The recording of the detector phase of the above systems depends on one of two methods of operation. If each reaction of a nuclear radiation with the detector is recorded and signaled, the counting is in terms of a series of pulsed signals. One example of this type is the Geiger-Müller tube. In the other type of operation, the measured quantity is an average effect of all the interactions of the radiation with the detector unit. There is, therefore, no resolution of individual events, and the current output is proportional to the events occurring in the detector per unit time. An example of this type is the ionization-chamber instrument.

It is evident that the detector unit of any radiation-measuring device can record only those radiation events that occur within its sensitivity and within its volume. Further, the current read on any dial, no matter how designated or complicated, is only a result of the meter design and circuitry and the conditions of use. Interpretation and subsequent evaluation of the dial indication can be made only when all limitations and use restrictions have been observed. Conversion to units of radiation exposure can be done only within relatively narrow ranges of energy and radiation intensities and is valid only when compared with accepted calibration procedures.

It must be remembered that any instrument measures only the radiation effect on that instrument's detector, and extrapolation of results to other radiation, other volumes, and other exposures is most uncertain.

Dose Concepts

Adding materially to the difficulties in biologic dosimetry is the multiplicity of units of radiation measurement. In most cases the exact

radiation spectrum is not known, and the intensity cannot easily be related to the ionization effects observed. Additionally, the absorbed dose, which is of much greater concern to the biologist, is more closely related to the ionization than to the intensity of the radiation.

Exposure Dose

Although many different dose measurements and criteria have been suggested and used, the majority have been concerned with the measurement of the exposure dose. After much discussion and influence based on local usage, the exposure dose was finally defined in 1956 as follows: Exposure dose of X or gamma radiation at a certain place is a measure of the radiation that is based upon its ability to produce ionization. The unit of exposure dose of X or gamma radiation is the roentgen (r). One roentgen is an exposure dose of X or gamma radiation such that the associated corpuscular emission per 0.001293 g of air produces, in air, ions carrying 1 electrostatic unit (esu) of quantity of electricity of either sign. Exposure-dose rate is the exposure dose per unit time. The unit of exposure-dose rate is the roentgen per unit time.[2]

Thus, the roentgen is a unit of exposure dose and represents a certain amount of X or gamma radiation that will produce in air a certain amount of ionization. Confusing the use of this definition is the fact that the term dose was sometimes used to refer to what is now designated exposure dose and sometimes to what is now called the absorbed dose. Depending upon the writer, and the context, and often indeterminable, the phrase "dose of *x* roentgens" could mean either "an exposure dose of *x* roentgens" or "an absorbed dose that would be expected from placing the material in an exposure of *x* roentgens."

Also, further confusion was added by the necessity for roentgen-equivalent dose measurements. Since the roentgen was defined only in terms of X or gamma radiation, the roentgen-equivalent-physical (rep) was used to equate the ionization capability of particulate radiation. To relate energy transfer and resulting ionization in media other than in air as defined for the roentgen, the roentgen-equivalent-biological (reb) and the roentgen-equivalent-mammal (or man) (rem) came into common usage, in addition to the rep.

Absorbed Dose

Progress in the right direction was made by attempts to evaluate and relate the amount of energy absorbed per volume of absorber, and thus the gram-roentgen concept in its various forms was initiated [7].

[2] *Ibid.*

Thus the r (air) represented an energy absorption of approximately 83 ergs/g, while the rep (tissue) involved nearly 93 ergs/g. The finally derived expression for the locally absorbed energy is the radiation-absorbed-dose (rad), which is defined as follows: The absorbed dose of any ionizing radiation is the energy imparted to matter by ionizing particles per unit mass of irradiated material at the place of interest. The unit of absorbed dose is the rad. One rad is 100 ergs/g.[3] Thus the absorbed dose is not a property of the radiation beam alone; it results from the interaction of the radiation with matter and therefore depends upon the properties of the particular material involved as well as upon the characteristics of the radiation.

Biologic Dosimetry

In spite of the extensive theoretical and technical knowledge in the electrophysicochemical aspects of radiation dosimetry, the application of these methods to the biologic systems remains largely empirical. One can hardly claim that biologic dosimetry is refined when, in radiation therapy, dosage is still recorded in many places as air dose. Even where tissue dose is recorded, the visual estimation of the depth of the lesion or tumor suffices to place the evaluation of the volume in the correct column in the depth-dose tables available. Even when localized by radiographic means, the midplane tumor dose is confidently recorded. In cases where an estimated tissue volume is considered, a mean dose is often used. With such lack of precision, biologic dosimetry has not really progressed very far beyond the earliest recognition and use of an erythema dose as therapy-machine calibration. But there is hope for early rectification and solution of this major problem. Many investigators, both in the medical and the allied biophysical sciences, are devoting great efforts to this study, and results are already promising.

The exposure to ionizing radiations can produce changes in a number of the physical properties of certain insulating solids. Many of these changes in the solid-state materials are of application to dosimetry. In itself the solid-state dosimeter is no more biological than the ionization chamber, but its capability for miniaturization and the direct utilization of its conductive properties (as with diodes) make it possible for these detector units to be placed at specific, predetermined positions within the biologic system. Series of these detectors can measure simultaneously the radiation levels and can give spectrum discrimination at many positions near and within an organ or other specific volume. The calculation of entrance, scatter, exit, and absorbed doses with such techniques far surpasses the type of estimation so often done now.

[3] *Ibid.*

However, the search for and development of a true biologic dosimeter continue. Of great value would be some substance, either from the body or completely compatible with it, that retains evidence of radiation exposure, that reacts in a graded manner to varied amounts of radiation, and that can be sampled conveniently. Some body fluids do show changes directly related to the radiation dose, and their application awaits, perhaps, only the refinement of a specific technique for isolation and recording of response [10].

Recent studies [10] with electron paramagnetic resonance measurements show some promise in free-radical detection and measurement in analogs of amino acid and proteinaceous material. The possibility of similar changes in protein-containing fluids of the body, such as in saliva or urine, is being studied.

The likelihood of metabolite alteration as the result of radiation exposure raises the hope that some of the products not normally excreted might be found in amounts related to radiation exposure. This sort of investigative survey continues to be extensive, but it requires an imaginative perspective that is often blinded by the details of meticulous technique.

BIBLIOGRAPHY

1. Cork, James: "Radioactivity and Nuclear Physics," 2d ed., D. Van Nostrand Company, Inc., Princeton, N.J., 1950.

2. Glasser, O., et al.: "Physical Foundations of Radiology," Paul B. Hoeber, Inc., New York, 1952.

3. Glasstone, Samuel: "Sourcebook on Atomic Energy," D. Van Nostrand Company, Inc., Princeton, N.J., 1950.

4. Glendenin, L. E.: Determination of the Energy of Beta Particles and Photons by Absorption, *Nucleonics*, vol. 2, p. 12, January, 1948.

5. Gray, L. H.: The Ionization Method of Measuring Neutron Energy, *Proc. Cambridge Phil. Soc.*, vol. 40, p. 72, 1944.

6. Kaplan, I.: "Nuclear Physics," Addison-Wesley Publishing Company, Reading, Mass., 1955.

7. Mayneord, W. V.: Energy Absorption of x- and Gamma-rays in Tissue, *Brit. J. Radiol.*, vol. 13, p. 235, 1957.

8. Price, William J.: "Nuclear Radiation Detection," p. 18, McGraw-Hill Book Company, Inc., New York, 1958.

9. Report of the International Commission on Radiological Units and Measurements (ICRU), *Natl. Bur. Standards* (*U.S.*) *Handbook* 62, 1957. [Also in *Natl. Bur. Standards* (*U.S.*) *Handbook* 78, 1961.]

10. U.S. Air Force, School of Aerospace Medicine, unpublished data.

11. White, George N.: "Principles of Radiation Dosimetry," p. 40, John Wiley & Sons, Inc., New York, 1959.

Chapter 7

THE ACUTE RADIATION HAZARD

*Dean W. Williams**

THE PRIMARY CAUSE OF CONCERN ABOUT HUMAN EXPOSURE TO IONIZING radiation is the threat of nuclear explosions or disasters. The hazards associated with the improper use of medical and dental X rays are important but controllable. Similarly, the normal background radiations of cosmic origin, normally occurring soil isotopes, television tubes, radium wristwatch dials, and the like, are of minor consequence.

An air burst of a nuclear weapon releases approximately 5 per cent of its total energy in the form of ionizing radiation that, outside of the fireball, consists primarily of neutrons and gamma rays. The duration of this initial radiation pulse is about that of the fireball. The balance of the weapon's energy is released in the form of blast and shock waves, thermal radiation, and residual radiation. The initial pulse of neutrons and gamma rays has a penetrating range through air and matter that depends upon the yield of the weapon and the density of the matter with which the radiation interacts. For example, from a 1-megaton air burst the lethal radius of initial radiation alone would be about 1¼ miles [15]. A ground burst releases this same pulse of neutrons and gamma rays and, in addition, contaminates the ground materials that are drawn up into the fireball. This radioactive ground material, along with bomb residues, constitutes the fallout material, which emits beta particles, gamma rays, and, to a lesser extent, alpha particles. The alpha, beta, and gamma radiation from the debris and the neutron and gamma radiation from the initial pulse constitute the major short-term radiological threat from nuclear weapons. Although there is a long-term radiological threat from these radiations, this chapter is concerned solely with the acute biological effects attributable to the initial high-intensity radiations and the immediate—occurring in a month or less—local fallout.

Alpha, beta, gamma, and neutron radiation differ in the degree of biological damage that they inflict. Alpha and beta particles have little penetrating ability, and their energy is rapidly attenuated in such low-density materials as air, wood, and human tissue. For example, the tissue-penetrating range of a beta particle is from 2 to 4 mm; the range is even

* The MITRE Corporation, Bedford, Mass.

shorter for an alpha particle. Therefore, protection against the alpha and beta portion of fallout can be readily achieved by preventing its contact with the skin and by preventing its inhalation or ingestion.[1] On the other hand, neutrons and gamma radiation can travel considerable distances through air and ordinary building materials and penetrate deep into the tissues of the body. These radiations are not easily attenuated and can cause considerable ionization in their passage through human tissues [15].

A number of terms have evolved in the radiobiology literature in an attempt to reconcile (1) the difference between an exposure dosage and an absorbed dosage and (2) the differences in biological effect resulting from absorption of the various forms of radiation. In the previous chapter, the terms roentgen, roentgen-equivalent-physical (rep), roentgen-equivalent-mammal (rem), roentgen-equivalent-biological (reb), and rad were defined. Each of these terms has advantages and disadvantages in the description of biological effects. Another radiation-effect term, effective biological dose (EBD), also may prove useful, because its computation includes additional detailed information. Specifically, computation of the effective biological dose requires precise description of the type of radiation, the rate of exposure, the duration of exposure, the decay rate of the radiation(s), the rate of repair of the involved tissues, the amount of irreparable damage, and so forth. It is an involved concept, but, if standardized, it will be particularly useful in describing the long-term effects of fallout exposure on biological systems [8].

In order to appreciate the intensity of radiation exposure that is dealt with later in this chapter, it is worthwhile to examine the magnitude of the dosages normally encountered from background sources and X-ray machines. Background radiation from cosmic rays at sea level and natural soil isotopes yields about 0.15 roentgen (r) per individual per year [15]. Efficient chest X-ray machines may deliver as little as 20 milliroentgens (mr) per exposure, whereas inefficient or improperly used equipment may deliver doses of 1 r or more. Similarly, some dental X rays may deliver as much as 300 r to the jaw, whereas better X-ray pictures can be obtained by using only 1.5 r with properly adjusted equipment [13].

The National Committee on Radiation Protection has made recommendations on the permissible doses from external sources of ionizing radiation for personnel occupationally involved with radiation [26]. These recommendations on the maximum permissible dose (MPD) define a permissible dose as that amount of ionizing radiation that is not expected to cause appreciable bodily injury to a person at any time

[1] Beta radiation that does contact the skin will cause first- and second-degree burns if the dosages are of the order of 5,000 rad or more [5].

during his lifetime. The underlying philosophy of these recommendations is that any exposure to ionizing radiation above that of the normal background increases the probability of somatic or genetic injury. Consequently, the maximum permissible dosage is an estimate of *acceptable* risk based on the limits of present knowledge. The basic recommendations are as follows: The accumulated whole-body dose for radiation workers is not to exceed 5 rem multiplied by the number of years beyond age eighteen, and the dose in any 13 consecutive weeks shall not exceed 3 rem. Thus, an individual thirty years of age shall accumulate not more than $(30 - 18) \times 5$ rem, or 60 rem, during his lifetime. The basic permissible weekly dose to either the blood-forming organs or the gonads is not to exceed 300 mr.[2]

THE CELLULAR EFFECTS OF IONIZING RADIATION

There are three possible ways in which beta, gamma, X-ray, and neutron radiations dissipate their energy when they strike target materials. Upon encountering a target atom, the radiation may cause the ejection of an electron such that ions are created. A separation distance of 50 to 100 μ often makes the immediate recombination of these ions impossible, and they may exist for finite periods of time before being neutralized. During these periods, the ions may combine with biochemical compounds and thus alter the character of the compound(s). If the radiation energy is not adequate to cause the ejection of an electron, it may shift the position of electrons in their orbital rings, thus changing the atom's reactivity [29].

When a neutron of high kinetic energy collides with a normal nucleus, the neutron transfers some of its energy to the nucleus and thus leaves the nucleus in a higher energy state. If a high-energy neutron strikes a small atom such as hydrogen, the hydrogen nucleus may be freed from its electron and move off as a high-energy proton. This proton in turn collides with other atoms and causes ionization of these secondary targets. If the neutron does not eject the nucleus but leaves it in a higher state of energy, the high-energy target nucleus may then dissipate its induced energy by emitting gamma rays. The gamma rays in turn may cause ionization of other atoms or compounds. Thus neutron radiation produces ionization indirectly, whereas gamma rays directly cause ionization.

Experimentation has shown that the exposure of a cell to 400 r produces less than one-millionth of the energy normally expended by the

[2] For additional literature on permissible doses of radiation see References 3 and 22 in the Bibliography at the end of the chapter.

cell in a 24-hr period. Exposure to 1,000 r might modify only 1 in 10 million molecules in the average-sized cell. It is presumed, therefore, that ionization must modify a key function of the cell for the radiation to exert an effect that is profound in comparison with the amount of the energy change [16]. There are two ways in which the ionization of intracellular biochemical compounds may change their chemical character. First, ionization may disrupt the chemical composition so that its reactivity is stopped and it is denatured into its basic biochemical constituents. Second, the ionization of water molecules may cause them to undergo a series of chemical transformations that result in the formation of highly reactive radicals. These radicals (OH, O_2H, and H_2O_2) may combine with the long molecular chains and in turn may disrupt the molecule's reactivity in enzymatic activity. There is evidence that the inhibition of key enzymatic function by radiation is the indirect action of irradiated water molecules rather than the result of direct collision with the enzyme molecule itself [4].[3]

It has been proposed that the presence of these oxidizing radicals within the cytoplasm of the cell is relatively harmless because bio-oxidations normally occur in the cytoplasm. However, the formation of these radicals within the nucleoplasm is harmful because most of the metabolism within the nucleoplasm is of an anaerobic nature and is extremely sensitive to the presence of the oxidizing radicals. In that the metabolic character of genetic activity within the nucleoplasm is highly specialized, this notion of nuclear sensitivity appears to have merit.

THE EFFECT OF IONIZING RADIATIONS ON MAMMALIAN TISSUES AND ORGANS

Although the damaging effects of radiation originate at the cellular level, different tissues and organs within the body show considerable variation in their response to ionizing radiation. First, rapidly growing cells are more radiosensitive than stable mature cells. Second, the maintenance and posture of resistant tissues and organs are intimately related to the condition of radiosensitive supportive tissues such as the vascular system. Therefore, major tissues and organs may be discussed in order of their relative radiosensitivity.

1. The blood-forming, or hematopoietic, system that resides in bone marrow is the most radiosensitive organ system in the body. Since both

[3] Chromosomes, for example, in which enzymatic activity is high, appear to be highly sensitive to ionization and will reproduce abnormal chromosomes during the mitotic cellular division. These chromosomal aberrations may be fatal to cells when they initiate their next division process [29].

direct and indirect destructive effects may be mediated through the vascular system, subacute radiation exposures, which may cause no subjective effects whatsoever, can often be detected by examination of the white blood cells. After a median-lethal whole-body dose (about 500 r) of gamma rays, the lymphocyte count begins dropping immediately and reaches its ebb within 3 to 6 days. The granulocyte count begins falling off in about 3 days, the platelets in about 9 days, and the red blood cells in about 12 days. There is a hemorrhagic tendency when the platelet count decreases and a reduced capability to resist bacterial invasion when the total count of white blood cells drops precipitously [19]. Therefore, symptoms directly related to changes in the blood constituents will be slow in onset, appearing perhaps in several weeks. The reason for this is that, whereas body cells, blood cells, or other types may have their capabilities impaired by radiation, the lethal cellular effect may not be encountered until the cell begins its mitotic division. Also, it takes some time for bacterial pathogens to penetrate injured tissue and begin proliferating in the blood stream.

2. The gastrointestinal tract is also relatively radiosensitive; as an organ system it is second in sensitivity to the blood-forming system. Depending upon the severity of the dosage, there is a range of direct effects on the gastrointestinal tract. At low levels of direct-radiation insult, the general physiological function of the gastrointestinal system is disrupted. In time, ulcers may appear in the tract that eventually cause direct hemorrhage. As the radiation insult is increased, the secretion of gastric juices is halted and at exposures greater than 1,000 r the permeability of the gut is so severely altered that body fluids are lost into the gastrointestinal tract [5].

Somatic homeostasis is highly responsive to alterations of the gut. Irradiation of the gut induces a feeling of malaise and nausea that may in turn induce vomiting and/or diarrhea. If these symptoms are acute and large amounts of fluid and electrolyte are lost, the blood's acid-base balance may be disturbed and the blood chlorides reduced. Drastic reductions in blood chlorides can result in neurological changes as severe as tetany.

3. The endocrine system, with the exception of reproductive tissue in the gonads, is relatively radioresistant. Supralethal radiation dosages must be absorbed before the endocrines are directly involved [19]. The adrenals, although not sensitive to radiation, do show evidence of increased activity after whole-body exposure. In fact, this adrenal response may be essential to the individual's survival to dosages in the lethal range [11]. Whether the adrenals respond to toxins released into the systemic circulation from injured cells or to a general systemic stress reaction is not clearly understood [12, 21].

4. The central and peripheral nervous system is perhaps the most radioresistant organ system in the body. Supralethal doses (5,000 r) are required to elicit gross neurological changes, but some functional changes, such as decreased excitability and imbalances between excitation and inhibition, have been observed at near-lethal doses (around 700 r) [19].

The nervous tissue of the eye enjoys a similar high degree of radioresistance. Destruction of the rods requires 1,700 to 2,000 r, and destruction of the cones requires 10,000 to 30,000 r [14]. The lens, cornea, and conjunctiva are all more sensitive than the retina. The approximate radiation threshold for lens-opacity effect (cataracts), for example, is about 500 r [7], which is about median lethal for human beings.[4]

5. Skin and muscle are also radioresistant [19]. Supralethal dosages must be encountered before the radiation causes direct effects.

VARIABLES ASSOCIATED WITH RADIATION EXPOSURE

An instantaneous whole-body exposure may be as small as that from normal background radiation or as large as that which is lethal for 100 per cent of the exposed personnel. Within this range there are not enough data to specify the exact somatic or genetic effects for each increment of radiation. However, one finding is clear, that the probability of occurrence of manifest symptoms increases as the dose increases. Table 7-1 gives a general expression of the expected effects of acute whole-body irradiation. It should be recognized that the descriptions of the higher dose levels will be more accurate clinically than those of the lower end.

In Table 7-1, a range for the median-lethal dose is specified as being that between 400 and 500 r. The concept of median-lethal, or LD_{50} (lethal dose for 50 per cent of an exposed population), has been generally used as a scaling concept. It gives no individual information, but it is a convenient reference, defining a midway point in the total range of gross effects. However, in regard to radiation exposure, the 50 per cent population that does survive should not be presumed to be unaffected by the radiation. The term $LD_{50/60}$ is used to specify 50 per cent mortality in 60 days. It has been suggested that this is a more convenient designation for human-population exposure than the simple LD_{50}. There is disagreement as to a specific LD_{50} for human beings; the range is 350 to 700 r.[5] It should be noted that the LD_{50} is not the same for the initial

[4] Experimental ocular lesions in animals can be produced with exposures around 2,000 r. These lesions are characterized by conjunctival congestion, swelling and narrowing of the pupils, and retinal edema. This damage, however, is due primarily to vascular irritation and breakdown [6].

[5] For individual estimates see References 15, p. 472, and 5, pp. 259–267, 269–273, 628.

radiation pulse at the fireball stage as it is for short-range fallout. An LD_{50} as low as 375 r has been estimated for fallout exposure because the radiation source would be entering the body from many directions, whereas an LD_{50} of 650 r might be more accurate for exposure to the initial pulse, which is unidirectional [23].

TABLE 7-1 *Expected Effects of Acute Whole-body Radiation Doses*

Acute dose*, r	Probable effect
0–50	No obvious effect, except possibly minor blood changes
80–120	Vomiting and nausea for about 1 day in 5 to 10 per cent of exposed personnel. Fatigue but no serious disability
130–170	Vomiting and nausea for about 1 day, followed by other symptoms of radiation sickness in about 25 per cent of personnel. No deaths anticipated
180–220	Vomiting and nausea for about 1 day, followed by other symptoms of radiation sickness in about 50 per cent of personnel. No deaths anticipated
270–330	Vomiting and nausea in nearly all personnel on first day, followed by other symptoms of radiation sickness. About 20 per cent deaths within 2 to 6 weeks after exposure; survivors convalescent for about 3 months
400–500	Vomiting and nausea in all personnel on first day, followed by other symptoms of radiation sickness. About 50 per cent deaths within 1 month; survivors convalescent for about 6 months
550–750	Vomiting and nausea in all personnel within 4 hr from exposure, followed by other symptoms of radiation sickness. Up to 100 per cent deaths; few survivors convalescent for about 6 months
1,000	Vomiting and nausea in all personnel within 1 to 2 hr. Probably no survivors from radiation sickness
5,000	Incapacitation almost immediately. All personnel will be fatalities within a week

* Received in less than 1 week.

SOURCE: S. Glasstone (ed.), "The Effects of Nuclear Weapons," U.S. Department of Defense, Washington, June, 1957.

The second variable associated with radiation-exposure effects is the rate of exposure. Although 500 r instantaneous exposure might be fatal to 50 per cent of the exposed individuals, this amount received over a period of 10 years would probably serve only to shorten the life span [14]. It is estimated that for adults receiving more than 100 r the life span is reduced by about 10 days for each additional roentgen [5].

A third exposure factor is the percentage of the body area that is irradiated. Whole-body instantaneous exposure of 500 r could be fatal to 50 per cent of the involved personnel, but this same dose might be used as medical treatment for a local malignancy. In fact, doses up to 10,000 r have been delivered to small tumors with no general reaction, and 400-r doses may be delivered to body areas of 20 by 20 cm and cause only

transient radiation illness [28] and an "acceptable" shortening of life span.

Another critical factor is the portion of the body that is irradiated. If certain portions of the body are shielded, the tolerance for lethal doses may be substantially extended. In animal studies, for example, when the skull, vertebral column, or pelvis was shielded, the experimenters could double the median-lethal dose. In fact, shielding only three or four vertebrae extended the period of survival. Head shielding has been shown to reduce radiation effects of the lining tissue of the alimentary canal. Preserving the integrity of the alimentary canal thus extended the survivability of the protected animals [1]. Since the upper abdomen is a highly radiosensitive area of the body [12], shielding the trunk as well as the head would offer even further protection for short-term exposures to radiation. Shielding the adrenals also has a short-term effect on survivability [10].

The physical condition of the subject is another factor that influences survivability (at least in the median-lethal range). A subject in poor physical condition is more apt to be sensitive to the indirect effects of radiation, such as the induced anemia, the hemorrhagic tendency, and the reduced defense capability of the white blood cells. Similarly, a subject exhibiting the physiological stress reaction is more radiosensitive than the resting animal [12, 14]. In contrast, having alcohol in the system seems to extend one's tolerance to radiation [15].

THE RADIATION SYNDROME

Observations of radiation effects have come from four main sources: (1) occupational accidents, (2) the use of isotopes and X rays for radiotherapy, (3) nuclear detonations, and (4) animal experiments. The occupational accidents include malignancies that resulted from overexposure to radioactive elements and X rays in the early days of nuclear research and, more recently, from incidents, such as the Los Alamos incident of 1945, in which research personnel were exposed to chain reactions in a nuclear reactor. When it was found that radiation has an adverse effect on rapidly reproducing cells, scientists began using X rays and isotopes as therapy for both malignant and nonmalignant cellular proliferations. This radiotherapy has given considerable information on radiation effects. After the atomic detonations over Hiroshima and Nagasaki, a study was initiated to determine the effects of a nuclear-weapon burst [17]. Information on the effects of fallout on human beings has been gleaned from a nuclear-weapon test in which fallout unex-

pectedly drifted over inhabited islands. Finally, a considerable number of data have been collected on the effects of radiation on animals.

The amount of reliable information about radiation effects on human beings is limited in quantity. Although there were a large number of victims in Japan, the dosimetry could not be accurately determined and therefore symptoms and aftereffects could not be reliably correlated with dose. In the few clinical cases that are reviewed later in this chapter the dosimetry was reliably determined.

When the possibility of nuclear attack is considered, civilian and military defense planners and physicians may find it convenient or necessary to group radiation victims into categories based upon expected symptoms. As the intensity of dosage is increased, its damaging and/or lethal effects are inflicted on different body-organ systems. Therefore, for clinical convenience in the treatment of large numbers of patients, the following four categories have been proposed [5]:

1. No obvious disease. The portion of personnel exposed to as much as 200 r may have transient, intermittent nausea up to the third day after irradiation. The degree of expected vomiting is dependent upon the dose, individual radiosensitivity, amount of stomach contents, and psychological disposition.[6] These victims are presumed to require no specific therapeutic measures or hospitalization.

2. Hematopoietic syndrome. The second category of radiation illness is characterized by sufficient radiation, 200 to 1,000 r, to involve or "insult" the blood-forming system. The clinical symptoms for this category may include any of the following: nausea and vomiting up to several hours after irradiation, fever late in the illness probably because of secondary infection, and diarrhea at the same time as vomiting, although this last symptom is not a certainty.

Increased bleeding tendencies appear later in the course of the illness, within approximately 2 to 3 weeks. Prominent changes in the blood stream include: a decrease in white-blood-cell count and a decreased lymphocyte count soon after exposure. This symptom, if it can be analyzed, may give one of the more accurate assessments of both exposure and prognosis. Specifically, a drop from the normal lymphocyte count of 1,000 to 3,000 per cubic centimeter to 500 per cubic centimeter probably puts the patient in the fatal category. The platelet count decreases and is related to the bleeding tendency later in the illness. Secondary infections may appear later in the illness because of the decreased infection defenses of the blood stream and because of the loss of integrity of the gut-lining tissues. Death will occur in approximately 50 per cent of the victims in

[6] Japanese physicians attributed much of the vomiting at Nagasaki and Hiroshima to psychological conditions [9].

the dosage range of 350 to 700 r and in 50 to 100 per cent of the victims in the range of 700 to 1,000 r.

This is the largest and most difficult category of patients because diagnosis is difficult, the symptoms may be protracted, and the range of effects is from no immediate effect to death. Large quantities of therapeutic agents will be required, and the patients will require care for long periods of time. There is no single set of therapeutic measures for patients in this category other than "good patient care," consisting of rest, antiemetics (to control vomiting), nutritional supplementation, water- and salt-balance maintenance, antibiotics, transfusions, and possibly, bone-marrow therapy.

3. The gastrointestinal syndrome. The third category of illness is for those receiving between 1,000 and 5,000 r, a dose involving both the blood-forming system and the gastrointestinal tract. The symptoms will be rapid in onset, severe, and of short duration, ending almost always in death within 2 weeks. The principal symptoms are nausea, vomiting, fever, diarrhea, and general debilitation. After several days of these initial symptoms, a short cessation of symptoms may be enjoyed, but the symptoms will reappear prior to death. As a class, these individuals will be easier to diagnose, and the degree of incapacitation will preclude the possibility of their being able to participate in normal activities.

4. The central-nervous-system syndrome. The fourth category of radiation according to organ-system insult is characterized by direct involvement of the central nervous system. Here, the required whole-body dosage is greater than 5,000 r. These patients can be expected to be completely incapacitated in minutes after exposure, and death will occur in hours.

Since it is unlikely that civil-defense and medical personnel will have instruments capable of performing body dosimetry, they will be forced to make decisions as to the care and disposition of patients on the basis of obvious symptoms. On this basis, radiation victims might be sorted into three groups.

1. Survival improbable. If vomiting occurs promptly after exposure, continues over several hours, and is followed by diarrhea, fever, and prostration, the prognosis is grave. Death can be expected within a week or two.

2. Survival possible. Nausea and vomiting may occur soon after exposure but will be of short duration. After a day or two of mild symptoms, the patient will enter a period of relative well-being, although weakness may still prevail. The longer this period of well-being, the greater the probability of survival. After the latent period of 2 to 3 weeks, new symptoms will appear, primarily because of changes in the blood stream. Therapy may be directed at bolstering the lagging bacterial defense

capability, preventing hemorrhaging from ulcerated tissue, maintaining proper electrolyte balance, and supplementing the diet to provide for nutritional losses in the ailing gut. The early course of the disease may be traced by taking a total white-blood-cell count a week after exposure. If the total white-blood-cell count is below 800 per cubic centimeter, prognosis is grim, whereas if it is greater than 1,500 per cubic centimeter, prognosis is encouraging. This group will be highly responsive to medical care [24].

3. Survival probable. Victims experiencing only transient nausea and malaise and a leveling off of depressed lymphocyte count within 24 to 48 hr may be expected to survive with a minimum of treatment [23].

CLINICAL CASE HISTORIES OF IRRADIATED PERSONNEL

One example of radiation exposure in which reasonably accurate dosimetry was possible is a fallout incident that occurred in 1954. Following a test detonation in the Pacific proving grounds, fallout settled on the island of Rongelap. Sixty-four native inhabitants were directly exposed to the fallout and were evacuated for medical treatment and rehabilitation.

It was estimated that the maximum absorption for some victims over a 2-day period was 175 r of penetrating gamma radiation and possibly as much as 5,000 r of beta radiation confined to the skin. Two-thirds of these people were nauseous during the first 2 days, and a smaller fraction experienced occasional vomiting and diarrhea. Two days after direct exposure to the beta radiation, the victims experienced transitory itching and burning of the skin and some lachrymation. Two weeks later, along with partial epilation, skin lesions developed on the neck, at the junction of arm and torso, and between the toes. The localized skin lesions were superficial (in depth), much like severe sunburn. The patients were not incapacitated by the skin lesions, and examination of the burn scars 4 years after exposure showed no further breakdown. No spectacular or statistically significant aftereffects of the radiation could be detected in a thorough 5-year study [5].

In the Los Alamos incident previously noted there were cases falling into each of the four disease categories. Here the radiations were fast neutrons, hard (penetrating) gamma rays, and soft (low penetrating ability) X rays. The whole-body dosimetry was ascertained with reasonable accuracy, but the dose to the hands of individuals touching the reactor could only be approximated. This approximation is noted only for academic purposes considering the severity of exposure.

In the first disease category, no obvious disease, there were seven

victims, ranging from one who received as much as 186 r soft X rays and 10.7 r gamma rays to one who received as little as 31 r soft X rays and 1 r gamma rays. None of these victims experienced gastrointestinal symptoms, nor did they report any subjective complaints. They resumed normal activity after a period of prescribed rest, and an examination 4 years later revealed no changes induced directly or indirectly by the radiation.

In the second disease category, hematopoietic syndrome, with doses of 200 to 1,000 r, there was one victim irradiated by 390 r of 80-kv soft X rays and 26.4 r of gamma rays. The patient vomited once several hours after exposure; after this there were no gastrointestinal disturbances. For several days he reported feeling weak and tired, and on the sixth day his temperature rose but soon returned to normal. In 15 days he was released from the hospital and in 10 weeks regained complete physical endurance. For a period of 4 years he was beset with a transient low sperm count, but 58 months later his wife had a normal child. He developed an incipient cataract in the lens of the right eye that reduced visual acuity. At the time of the last examination this was the only apparent residual effect of the radiation.

The following two victims might be placed in either the third or the fourth category, for, although their whole-body dosages were probably less than 1,000 r, their hands received several thousand roentgens. In one case the patient was exposed to 480 r soft X rays, 100 r gamma rays, whole body, and as much as 40,000 r soft X rays on his hands. This patient was in distress and was prostrated for 24 hr following exposure, after which time he was alert. On the fifth day he developed a fever, and he declined slowly until his death at 25 days. In the second case the patient was exposed to whole-body radiation of 1,930 r soft X rays, 114 r gamma rays, and up to 30,000 r soft X rays on his hands. He was ill within 1 hr after exposure; this was followed by a 5-day period of good general condition. On the fifth day the white-blood-cell count dropped, and on the sixth day the patient's fever rose and his pulse rate increased until his death at 9 days [16].

BEHAVIORAL EFFECTS OF IONIZING RADIATION

Beginning in 1951, an Air University research group conducted a study of 263 patients receiving radiation therapy for systemic neoplasms [5, 25]. In addition to the clinical observations made to determine the course of the malignancy after radiation, the researchers administered psychomotor-performance tests consisting of the two-hand coordinator and the rotary-pursuit tests. The dosages were by 15-, 25-, and 50-r in-

crements, with total doses of 25 to 200 r. It was concluded from observations of psychomotor skills that there was no evidence of psychomotor decrements among the irradiated individuals, whether the radiation was administered over a period of a few minutes or whether it was administered over a period of a day in five different fractions of dose. Also there was no clinical evidence of radiation effects.

Animal studies have been conducted to determine whether or not radiation exposures affect learning or retention [18, 27]. For example, monkeys receiving median or just-lethal doses were tested for acquisition, retention, and transfer of multiple discrimination problems immediately before, immediately after, and 150 days after exposure. No significant changes in score were observed, and the only reported performance deficit was a decrease in reaction time. In another study, rats were irradiated and put through an exhaustion swimming test [20]. Rats receiving 300 r (LD_{50} rats = 700 r) had slightly less endurance than did the controls, and those receiving 500 r showed significant performance decrements. Performance proficiency gradually decreased, reached its minimum during the third and fourth weeks after exposure, then returned to normal by the ninth week after exposure. This performance decrement was probably due to somatic malaise rather than to any direct effects on neuromotor functions.

One review of the research on the effects of radiation on behavior concludes that there are no demonstrable behavioral changes resulting from radiation [14]. When lethal dosages were used by the investigators, the animals succumbed to general systemic collapse without direct involvement of the central or peripheral nervous systems. Histological examination of the nervous tissue confirmed these observations. These conclusions on radiation behavior are consistent with evidence about radiation victims in Nagasaki and Hiroshima. Clinical data from the observation of 49 victims at about 1,000 m from ground zero (where the dose was greater than 1,000 rad [2]) reported only one case of visual disturbance late in the illness and one incident of delirium, also late in the illness, which was assumed to be due to the high fever in the terminal stages of the illness [17].

CONCLUSIONS

Although there is only a limited amount of information on human radiation effects, there is enough information from which to form some generalizations about the likely effects of radiation upon normal healthy individuals. Despite the fact that increasing levels of radiation mediate their effects through different organ systems and physiological functions,

the clinical history of irradiated personnel follows a fairly constant course of events in relation to dosage. The range of exposure from no observable effect to 100 per cent lethality is about the same for all mammals, that is, from 200 to 1,000 r. However, there are several conditions of exposure that reduce the effects of radiation. If some of these protective conditions, such as partial body shielding, are enhanced, survivability is significantly increased.

After a nuclear blast, radiation dosimetry would be difficult, if not impossible, to perform. Therefore, judgments as to the treatment and disposition of irradiated personnel will have to be based upon two criteria, (1) the observable behavior and condition of the victim and (2) limited clinical methods such as blood examination. There is no single reliable clinical symptom that characterizes the degree of radiation damage. Therefore, judgment as to individual care will have to be made on a moment-to-moment basis.

If the dosage is between 100 and 300 r, the blood-cell picture will change at about the time of the onset of the prodrome. During this initial period, individuals may be nauseous and ill for a length of time dependent upon the dosage and individual susceptibility to radiation. Those who do not show fixed symptom patterns may display transient symptoms that could prove disruptive to their normal activity.

As the dosage approaches the median-lethal range, between 300 and 500 r, the initial symptoms will be more prominent and of longer duration. Following the prodrome will be a temporary cessation of symptoms even though the patient may be somewhat weakened. After this period of remission, new symptoms will appear that are more closely related to the vulnerability of the hematopoietic system and its vitally dependent structures. Now the patient will enter a period of general systemic deterioration, in which secondary infections can flourish. This phase will debilitate victims for periods of weeks to months, depending upon individual variability.

Victims exposed to doses well within the median-lethal range, between 500 and 700 r, will be incapacitated for a longer time and will have a shorter remission period. Again, the return of symptoms will be related to the most vulnerable aspect of vascular-system support. Many of these victims will be responsive to long-term general medical care. However, at best, many will be left in a weakened condition for several months, and, at worst, some will succumb to infection, internal hemorrhage, or toxic septicemia.

Victims irradiated by 700 r or more may be considered as extensions of the previous class of patients. Many more will be sicker sooner and for longer periods of time. Most will die. When the dosage exposure is supralethal, individuals succumb immediately to systemic insult and to cumu-

lative organ-system collapse. The speed of onset, the duration, and the severity of symptoms will always give an accurate assessment of the extent of injury.

One important feature that was noted during a review of the literature on the radiation hazard was an absence of evidence of significant neuromuscular or neurologically induced behavioral disturbances directly due to radiation [14]. This phenomenon was true even for fatally irradiated victims. In fact, fatally exposed victims in Hiroshima and Nagasaki were able to engage in rescue activity during the remission period, which, incidentally, lasted in some instances for several weeks. What this suggests is that even a heavily irradiated person may maintain some performance capability until the biological insult precipitates disruptive feelings and symptoms. Whether or not an irradiated person will actually elect to function in a constructive capacity during symptom-free periods will depend upon emotional and social factors as well as biological damage.

The intent of this chapter is neither to exaggerate nor to minimize the radiation hazard. Rather, it is to emphasize the fact that performance decrements will be determined by a multiplicity of factors, both physical and mental. It would appear that healthy individuals can tolerate fairly high doses of radiation without dire consequence. The import of this fact is that, after a nuclear disaster, many individuals receiving moderate doses of radiation may have a good chance of survival and will even be able to contribute to the efforts of reconstitution. The long-term effects of radiation exposure are still unquantified and unqualified and in all probability will present some serious problems for future research.

BIBLIOGRAPHY

1. Allen, J. G., et al.: The Causes of Death from Total Body Irradiation, An Analysis of the Present Status after Fifteen Years of Study, *Ann. Surg.*, vol. 146, pp. 322–341, September, 1957.

2. Arakawa, E. T.: Dosimetry in Hiroshima and Nagasaki Atomic-bomb Survivors, *New Engl. J. Med.*, vol. 263, no. 10, pp. 488–493, Sept. 8, 1960.

3. Background Material for the Development of Radiation Protection Standards, *Fed. Radiation Council Staff Rept.* 1, May 13, 1960.

4. Barron, E. S., et al.: Inhibition of Enzymes by X-rays, *J. Gen. Physiol.*, vol. 32, pp. 537–552, 1949.

5. *Biological and Environmental Effects of Nuclear War, Hearings before Special Subcommittee of the Joint Committee on Atomic Energy*, 86th Congr., 1st Sess., June 22–26, 1959.

6. Cibis, P., et al.: Ocular Effects Produced by High Intensity X-radiation, *Am. Med. Assoc. Arch. Ophthal.*, vol. 53, pp. 651–663, 1955.

7. Claus, W. D. (ed.): "Radiation Biology and Medicine," Addison-Wesley Publishing Company, Reading, Mass., 1958.

8. Dixon, Harvey L.: Comparison of Effective Biological Doses for Three Alternative Recovery Rates, Project No. IV-2690, Stanford Research Institute, Menlo Park, Calif., December, 1959.

9. Dunham, C. L., et al.: Atomic Bomb Injury Due to Radiation, *J. Am. Med. Assoc.*, vol. 147, pp. 50–54, 1959.

10. Edelmann, A.: Adrenal Cortex and Survival of Rats after X-irradiation, *Fed. Proc.*, vol. 9, p. 36, 1950.

11. ——— and S. Katsh: Survival of Rats Adrenalectomized after X-irradiation, *Fed. Proc.*, vol. 10, p. 38, 1951.

12. Ellinger, F.: "Medical Radiation Biology," Charles C Thomas, Publisher, Springfield, Ill., 1957.

13. *Fallout from Nuclear Weapons Tests, Hearings before Special Subcommittee of the Joint Committee on Atomic Energy,* 86th Congr., vol. 2, May 5–8, 1959.

14. Furchtgott, E.: Behavioral Effects of Ionizing Radiation, *Psychol. Bull.*, vol. 53, no. 4, pp. 321–334, 1956.

15. Glasstone, S. (ed.): "The Effects of Nuclear Weapons," U.S. Department of Defense, Washington, June, 1957; revised ed., 1962.

16. Hempelmann, L. H., et al.: The Acute Radiation Syndrome: A Study of Nine Cases and a Review of the Problem, *Ann. Intern. Med.*, vol. 36, no. 2, pp. 279–510, February, 1952.

17. Hollingsworth, J. W.: Delayed Radiation Effects in Survivors of the Atomic Bombings: A Summary of the Findings of the Atomic Bomb, *New Engl. J. Med.*, vol. 263, no. 10, pp. 481–487, Sept. 8, 1960.

18. Kaplan, S. J., and G. Gentry: The Effect of Sublethal Dose of X-irradiation upon Transfer of Learning in Monkeys, *USAF School Aviation Med. Rept.* 4, Project No. 21-3501-0003, 1953.

19. Kay, F. DeWitt, Jr., and W. Hobson, Jr.: A Radiobiology Guide, Part II, *USAF WADC Tech. Rept.* 57-118 (II), May, 1959.

20. Kimeldorf, D. J., D. C. Jones, and T. J. Castanera: Effect of X-irradiation upon the Performance of Daily Exhaustive Exercise by the Rat, *Am. J. Physiol.*, vol. 174, pp. 331–335, 1953.

21. Leblond, C. P., and G. Segal: Differentiation between the Direct and Indirect Effects of Roentgen Rays upon the Organs of Normal and Adrenalectomized Rats, *Am. J. Roentgenol. Radium Therapy*, vol. 47, pp. 302–306, 1942.

22. Maximum Permissible Dose Burdens and Maximum Permissible Concentrations of Radionuclides in Air and in Water for Occupational Exposure, *U.S. Dept. Commerce Natl. Bur. Standards Handbook* 69, June 5, 1959.

23. *The Nature of Radioactive Fallout and Its Effect on Man, Hearings before Special Subcommittee on Radiation of the Joint Committee on Atomic Energy,* 85th Congr., pt. I, May 27–June 3, 1957.

24. Notes on Atomic Energy for Medical Officers Issued by the Royal Naval Medical School, Royal Naval Medical School, Farnsworth, 1956.

25. Payne, R. B.: Effects of Ionizing Radiation upon Human Psychomotor Skills, *USAF School Aviation Med. Rept.* 59–29, 1959.

26. Permissible Dose from External Sources of Ionizing Radiation, *U.S. Dept. Commerce Natl. Bur. Standards Handbook* 59, Sept. 4, 1954; revised, Apr. 15, 1958.

27. Rogers, C. M., et al.: Some Effects of Cumulative Doses of X-radiation upon Learning and Retention in the Rhesus Monkey, *USAF School Aviation Med. Rept.* 11, Project No. 21-3501-0003, 1953.

28. Tullis, J. L.: The Response of Tissue to Total Body Irradiation, *Am. J. Pathol.*, vol. 25, pp. 829–851, September, 1949.

29. Zirkle, R. E.: Relationship between Chemical and Biological Effects of Ionizing Radiation, *Radiology*, vol. 52, pp. 846–855, 1949.

Part C

HUMAN FACTORS IN WORKSPACE AND CLOTHING DESIGN

Chapter 8

THE BODY SIZES OF TOMORROW'S YOUNG MEN

Russell W. Newman[*]

WHEN WE SAY "SIZE" IN REFERENCE TO YOUNG MEN, WE ARE TALKING about something that is not constant—young men continue to grow taller year by year until they reach their maximum height. They also grow heavier and, therefore, broader and thicker, although this familiar phenomenon is more appropriate to older age groups than to young men.

There are two ways of looking at or for growth. One is the longitudinal approach, where one actually measures the stature of the same person year after year and the differences are called "growth." This is the way that parents check their children's successive increments in growth. Another technique is called the cross-sectional method, where one studies a cross section of a generation. This is the only method that can be used with a military population, the principal source of body-size data on American young men, since there is no successive measurement of subjects. The most simple and effective method for portraying and examining such cross-sectional data is by plotting age vs. stature and by then expressing the statures of any age group by an average or the statures of the entire group by a regression line.

An example of the cross-sectional method of examining growth in two military samples is given in Figure 8-1. One is a 25,000-man sample measured in 1946, the sample that has been used for most quartermaster clothing-data analyses. In order to make the data more comparable with material that will be presented later, only that portion of the sample between the ages of seventeen and twenty-seven is shown. The other group represented in this figure is a sample of about 4,000 men who were battlefield deaths in Korea during the Korean conflict. The average statures and the calculated lines of best fit for these averages are shown.

The point of emphasis in this figure is that the increase in stature with age up to the midtwenties is of the usual and expectable pattern—some men (in ever-decreasing numbers) continue to elongate until all

[*] Environmental Protection Research Division, U.S. Army Quartermaster Research and Engineering Command, Natick, Mass.

have reached their maximum adult stature. However, if we continue to plot average stature vs. age for slightly older groups, we find that average statures become shorter. This finding may be interpreted in either of two ways: first, that we are seeing the first indications of an actual shrinking

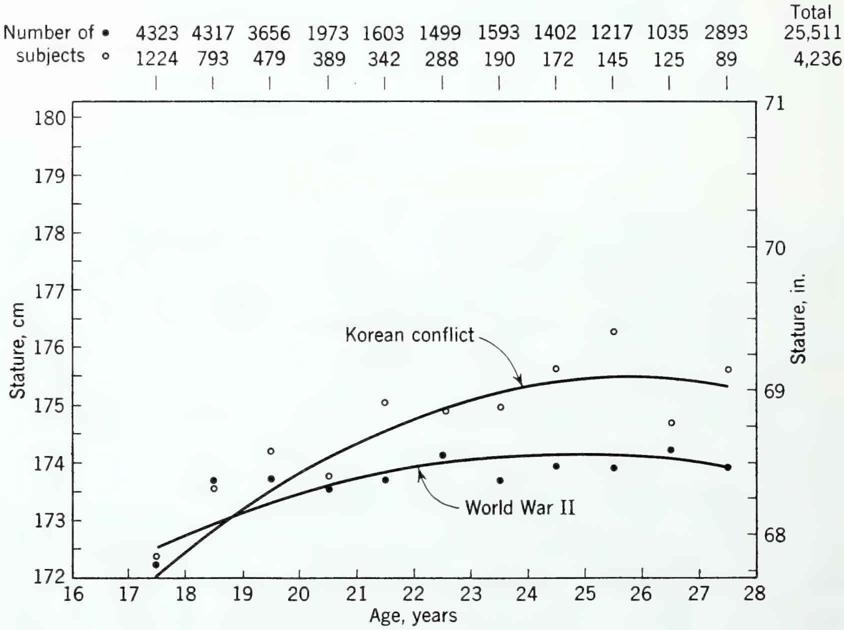


FIG. 8-1 *Mean stature by age group in World War II and Korean-conflict samples.*

of the individual by slow, steady compaction of the vertebral column, complicated by postural habits with increasing age; or, second, that we are seeing a "generational" change, since any plot that begins with the late teens and extends to the early forties may show a generalized "father-son" relation.

To facilitate war-dead identification, the Quartermaster Corps attempted to determine the stature increases that might occur in soldiers between the time that they entered the service and later death on the battlefield. To obtain supportive data on stature increase, the Quartermaster Corps asked three colleges to conduct a longitudinal-stature study on young men. The three institutions, the University of Michigan, the University of Washington, and Springfield College, each gathered four successive series of measurements (stature, sitting height, weight, and body fat) on a sample of college men over an 18-month period. On those men who remained in the particular college for the full 18 months, four sets of measurements were obtained at approximately 6-month intervals.

At the beginning of the study, each college had slightly over 800 subjects, but attrition reduced the total number to only slightly over 800 for all three institutions.

Reliability of the stature measurements was essential to the success of the study. When small increments in total height must be detected, it is important that known sources of error are eliminated or reduced to the minimum. Such sources of error include the posture of the subject, the inclination of the head, and the small but significant decrease in stature from morning to night. A last source of error, the unknown but possibly large differences among measuring devices, was minimized by constructing and furnishing to each of the three universities a special device, which is shown in Figure 8-2. This device consisted of an

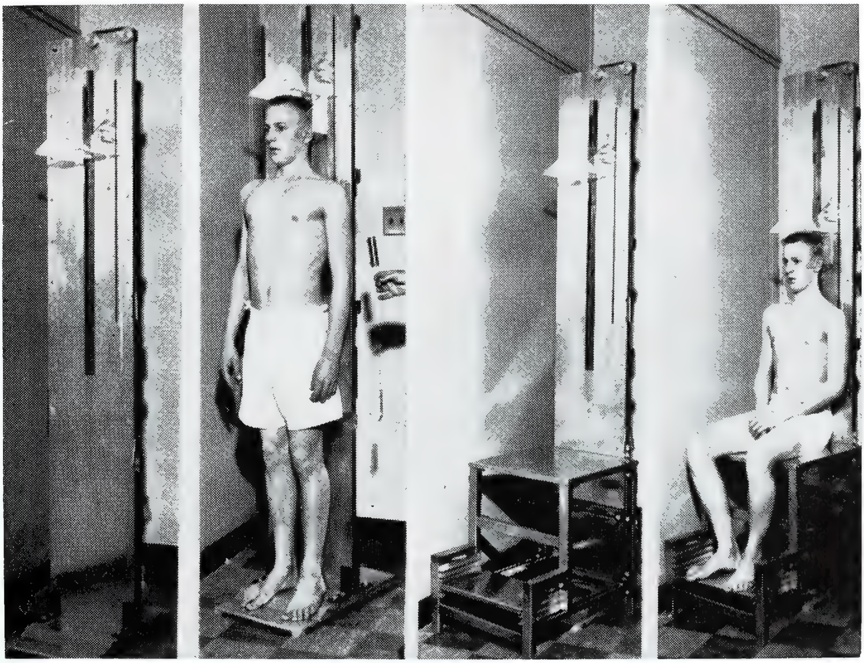


FIG. 8-2 *Apparatus and techniques used to measure stature and sitting height.*

aluminum upright on a steel base with a pulley-operated plastic box that rested on the top of the subject's head and aligned along a metric tape. For ensuring orientation in the eye-ear plane, a folding, adjustable triangle was fastened to the left side of the box. The box was counter-weighted to provide approximately 3 lb of weight on the subject's head during measurement. To achieve greatest reliability, a technique called "maximum-stretch" stature was used, where the subject elongates his body

by a full stretch while keeping his heels on the ground. So that subjects did not inadvertently apply pressure from the ball of the foot, it was arranged that the feet hung over a beveled portion of the base plate. Because of the care taken in this experimental arrangement, the statures recorded in this study may differ slightly from more conventional stature measurements.

Figure 8-3 shows the stature-vs.-age plot obtained from the first

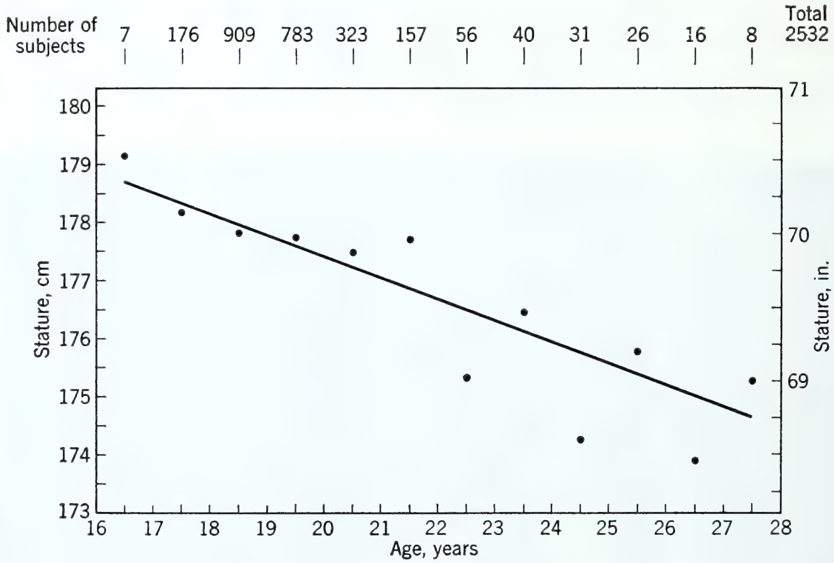


FIG. 8-3 *Initial mean stature by age group in the college study.*

measurements of the college sample, a total of 2,532 college men between the ages of sixteen and twenty-seven, drawn from New England, Middle Western, and Far Western institutions. The youngest and oldest age groups are represented by very small samples, but these small samples do not differ too widely from the regression line calculated for the entire sample. The principal feature of this figure is the remarkable negative slope of stature with age. Based on the regression line, the average-stature difference across the 11-year span is about 1½ in., which is an amazing difference for averages.

Figure 8-4 represents the data from the World War II, the Korean conflict, and college studies, plotted on the same graph for direct comparison. To place these data into proper perspective, the averages are plotted against year of birth, a method that reverses the curves from Figures 8-1 and 8-3 but does not change the data. The World War II

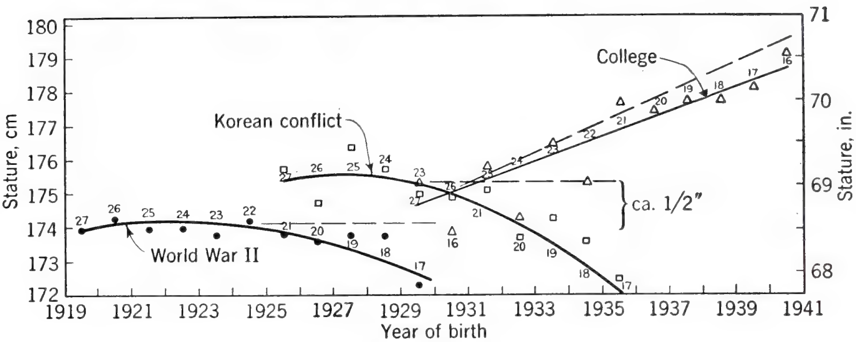


FIG. 8-4 *Stature on age by year of birth.*

data are shown in the lower left-hand corner of the figure, the Korean sample in the center, and the college sample in the upper right.

Some interesting results emerge from ordering the data in this manner. First, the World War II and the Korean-conflict teen-agers of the same age are about the same height. Dashed lines were added to the figure to indicate the minimum growth that was felt might be projected for these younger men in each sample in order to bring them up to the stature level of the older members of their respective samples. In addition, there is a year-of-birth overlap between the World War II and the Korean-conflict samples for the years 1925 to 1929. If the Korean-conflict twenty-seven- to twenty-three-year-olds born in the years 1925 to 1929 were truly representative, we may conclude that the teen-agers of World War II also born in 1925 to 1929 exceeded the limits represented by the dashed line for the World War II sample when they finally achieved adult stature. The same reasoning and conclusion may be applied to the overlap between the Korean-conflict and the college samples in the early 1930s. Even more striking is the finding that the college students born in the last half of the 1930s look as if they are part of a different population.

Before the significance of these data can be assessed, there is one additional pertinent result of the successive measurements in the college study that must be discussed. Originally stature growth was to be measured at 6-month intervals. At each of these intervals it would not have been surprising to find no stature increases, since it seemed possible that the younger subjects had already attained their final adult stature. However, 6-month increases in stature were found. Figure 8-5 shows the increases in those subjects who were measured yearly. (Only about half the original number of subjects were available for these measurements.) Their stature change was plotted by connecting the beginning and the 1-year average values for each age group. From this it is apparent that

perceptible growth occurred in each age group, in almost inverse proportion to age (the younger the age, the greater the growth). The 1-year stature growth is used here because it is easy to think in terms of 1 year's growth. The 6-month data are equally useful—there were merely not as

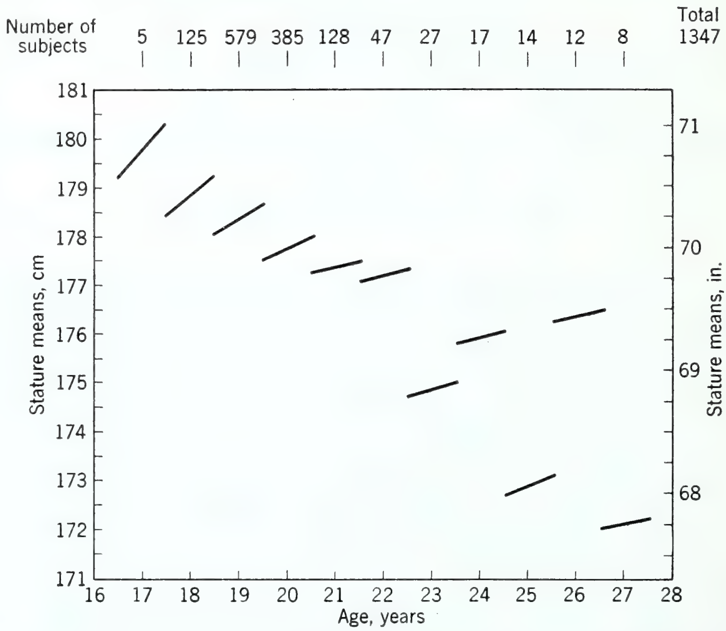


FIG. 8-5 *Twelve-month stature growth for each age group in the college study.*

marked changes. The 18-month material is also consistent but rather sparse in all but a few age groups.

Figure 8-6 shows the growth in the successive average statures for seventeen-, eighteen-, nineteen-, and twenty-year-olds at the start of the study. It is reasonable to suppose that the other age groups would show similar results if a larger sample were available.

It seems most probable that this college sample uniquely reflects the stature increases with age that child-growth studies have been reporting for over a decade. Unfortunately, child-growth studies that use the cross-sectional technique cannot predict how tall their subjects will become—they can only indicate how the subjects compare with other samples at similar ages. One question that had remained unanswered in the stature-age work on children was whether or not the young "giants" that were reported had simply achieved their adult stature very early. If this were the case, there would be no great increase in the adult stature of their generation but only a change in the maturation rate. However, the college

study clearly showed that these tall individuals continue to grow and that their final stature might well be much higher than that of their parents' generation.

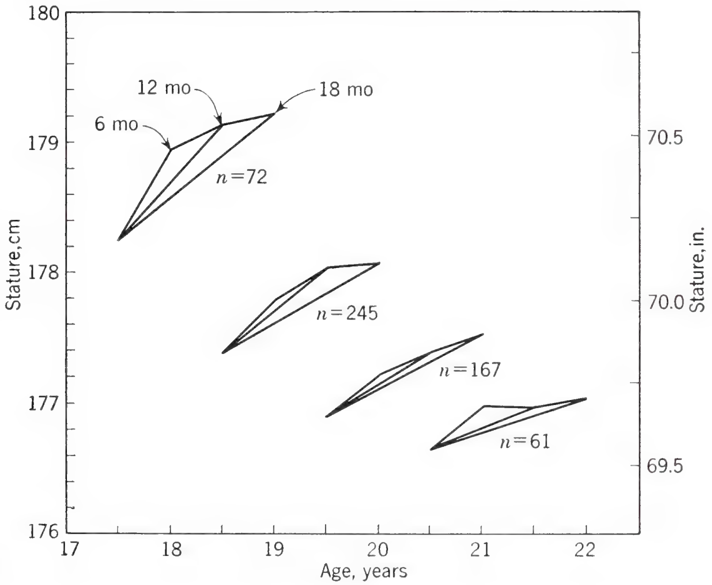


FIG. 8-6 *Six-month, twelve-month, and eighteen-month growth for the college men in each age group.*

It is worth mentioning that the college-sample subjects were born in the years between 1929 and 1940. Surely any economist would see clear and significant negative correlations between these age-group statures and the measures of national prosperity in those years. It would be unwise to infer a direct causal relation between the two, but if this stature increase is a result of better infant and child nutrition and hygiene in the 1930s, what should we expect from the children of the 1940s and 1950s when they reach adulthood?

We might suspect that in the near future the body sizes of our young men may change rapidly from what we are used to—perhaps even what we are prepared for. The stature changes that have been discussed here will be reflected in all those body dimensions functionally linked with stature, including the extremities and the head and face. Therefore, it is probable that all the workspace areas in man-machine systems that are even slightly restrictive for tall men will become increasingly unusable for future generations.

There is, of course, uncertainty in extrapolating from this small

college sample to a larger United States population. However, for the purposes of example only, let us assume that the future military population evidences the same age-stature relation as the college sample. For the example let us choose the youngest college group of which the sample size is reasonably large—the seventeen-year-olds. This group began in 1957 with an average stature of slightly over 5 ft 10 in. One year later those still in college averaged slightly over 5 ft 10½ in., and 6 months after that they had grown another small fraction of an inch. By 1960 this group was twenty years old, and it is fairly certain that their average stature was somewhere between 5 ft 10½ in. and 5 ft 11 in.

If this age group were to be inducted en masse into the Army, how would it fit the present clothing system? First, between 45 and 50 per cent of this group would be 5 ft 11 in. or taller. The present clothing system assigns all men 5 ft 11 in. or taller to the "long" category, and the published 1955 tariff figures on one example, the field coat or field jacket, expects only 14 per cent of the Army to require a long coat. Our age group would require more than three times that amount and would probably generate a fairly large requirement for new models in extra-long lengths. Simultaneously the requirement for short and regular lengths would be reduced. The present size system has the regular length neatly distributed over the World War II population stature average, but it seems inevitable that this distribution will become progressively less satisfactory when the children of the 1940s become the soldiers of tomorrow.

The college study did not give direct body-girth information, which might be translated into different girth-size requirements; however, it did provide information on body weight. Across the entire age range the college sample was on the average 12 lb heavier than the 1946 sample, and the age group that was used as an example, the college seventeen-year-olds, averaged 20 lb more than the comparable 1946 military age group. This would inevitably mean that the college age group would far exceed the present tariff for large sizes, which totals less than 7 per cent for the field jacket.

Let us now turn to the stature allowance of aircraft and tanks, an example of man-machine workspace that will be directly influenced by a change in population body length and sitting height. Here we are concerned not with the ordering and issuing of thousands of items such as clothing but with fitting a population into a few very expensive and intricate machine systems. The Air Force places a sitting-height upper limit of 38 in. on its potential pilots [1], because this is all the space that is available in the jet trainers. Army Ordnance uses a comparable figure, 38½ in., for guiding the design of driver's compartments in armored fighting vehicles [2]. Both these figures take into account clearance

allowances for headgear. These upper limits are reasonable and represent the 95th percentile of World War II and postwar military populations. How would the college subjects fare under such limits? Limits that exclude only 5 per cent of previous military populations will exclude between 15 and 20 per cent of the college subjects (slightly more of the teen-agers than those in the early twenties). This may not seem to be too serious in itself, but sitting height is only one of a number of cumulative selective criteria for pilots and tankers, the total effect of which may be very serious. Perhaps the potential magnitude of the problem can be shown by comparing the college sample with the 1961 anthropometric survey of Army flying personnel [3]. The sample of 500 officers and warrant officers was a fairly mature group (average age of thirty years) and, like all officer samples, was fairly tall (average stature of 5 ft $9\frac{1}{3}$ in.). This is a group that would draw heavily upon college graduates to replace its ranks. Although the flyers were tall, every college age group through age twenty-two exceeded their average, with the teen-agers much taller. Therefore, it seems reasonable to postulate that aircraft planned for use in 5 to 10 years had better be prepared to accommodate six-footers with the same facility that now exists for men of medium heights. The same can be said for tanks and other fighting vehicles. In fact, the whole defense establishment of tomorrow is going to have to provide more space in all man-machine systems for taller and larger operators.

For those researchers whose principal interests lie in basic science, this area of study may provide an unparalleled opportunity to observe our species undergoing rapid and unexpected biological change. To those more interested in applied research, it should be stressed that it seems unwise to allow such a biological phenomenon to appear without warning and without the best possible predictions as to its extent and duration. In short, the body-size problem is one that needs further investigation in a larger and more representative sample of young American men.

BIBLIOGRAPHY

1. Dempster, W. T., W. C. Gabel, and W. J. L. Felts: The Anthropometry of the Manual Work Space for the Seated Subject, *Am. J. Phys. Anthropol.*, vol. 17, no. 4, pp. 289–317, 1959.

2. Hansen, Alfred A. E.: Human Engineering Data and Factors, Military Vehicles, U.S. Army Ordnance Tank-Automotive Command, Detroit Arsenal, Center Line, Mich., 1958.

3. White, Robert M.: Anthropometry of Army Aviators, *USA Quartermaster Research Eng. Command Rept.* EP-150, 1961.

Chapter 9

COMPATIBILITY OF PERSONAL EQUIPMENT WITH WORKSPACE

*John L. Kobrick**

PERSONAL CLOTHING AND EQUIPMENT ARE IMPORTANT TOOLS FOR THE operator. The palm of a glove can help to provide a surer grip on a handle than would be the case with the bare hand. The sole of a boot allows greater pressure to be exerted on a pedal than would be possible with the bare foot. The conventional notion is that the clothed operator is an intact subassembly that functions in the total man-machine system. This assumption is used in the following discussion, and the problems that arise at this level of interaction are considered.[1]

One of the most common problems in the man-machine system is that of determining the relative size of system components necessary to accommodate the human being who will operate the equipment. The obvious solution is to measure the size of the operator and to build the system components so that the man will fit within the layout. Of course, there are always practical limitations to how far the workspace can be expanded to approximate an ideal solution. But let us assume for the moment that space is unlimited by other factors.

It is certainly recognized that human beings vary in size, but the implications of this fact are often not fully appreciated in system design. A very frequent error in the past has been the use of the arithmetic mean, or average, as the basis for sizing equipment. In a normally distributed population of body sizes—which is the best assumption for the design of equipment to be used by large numbers of people—the average becomes the midpoint of the distribution. This means that the equipment will accommodate only half the operators for whom it was intended,

* Engineering Psychology Laboratory, U.S. Army Quartermaster Research and Engineering Command, Natick, Mass.

[1] The Quartermaster Research and Engineering Command is concerned with developing optimum equipment and clothing for Army operators. In this role it has dealt with the problems inherent in making such equipment compatible with human functions and workspace requirements. Some of the Army equipment and its associated problems are used in this paper as examples for the more general topics that are discussed.

namely, those of average size or smaller. Large men will be unable to squeeze in or able to do so only with difficulty. A much more meaningful statistic for representing size requirements is the range, the spread encompassed by the largest and smallest values of a given measure. This kind of information gives a good picture of what kind of variation is involved. The practical spread to be considered for design purposes can be expressed conveniently by percentiles. For most sizing applications, the most suitable spread to consider has been found to be that encompassed by the 5th and 95th percentiles. This range tends to eliminate extreme variation and still takes account of 90 per cent of the people.

The sizing problem becomes more pronounced when the man is equipped with bulky protective clothing, such as high-altitude flying suits and uniforms used for cold-weather protection in arctic operations. Such gross gear increases body size significantly; furthermore, arctic clothing is semisoft, causing awkward bunching in tight spaces. The result is an unwieldy body form that must be allowed extra space to be accommodated adequately by surrounding hardware. Surely engineers and designers of hardware equipment used in conjunction with Quartermaster clothing must frequently wonder, and quite justifiably, why the size and bulkiness of arctic protective uniforms cannot be reduced. Unfortunately, there is not a great deal that can be done in this direction, for the uniform to a considerable extent depends for its insulation upon its looseness. This feature permits the isolation of dead air spaces as thermal barriers inside the suit and also minimizes binding at various body locations. Such binding would hasten body cooling by reducing blood circulation and thereby retarding the spread of heat through the body. In short, cold-weather uniforms of the present type just cannot be snug and remain effective.

The Army arctic uniform illustrates another important point for equipment design. Where bulky clothing ensembles are used by equipment operators and the size cannot be reduced, the use of nude-body size statistics for workspace sizing is a serious error. Such a practice is bound to limit the population of operators who will be able to use the equipment. Furthermore, if the mean for the nude body is the value used for a sizing index, the error will be even more serious. This will result in the exclusion of much more than half the population of operators for whom the equipment was designed. There is no need to describe the results of such errors' remaining undiscovered until the equipment goes into the field. For mitigating the many sizing difficulties in the use of the arctic uniform with other equipment, sizing information has been published. (See the appendix to this chapter for a discussion of this literature.)

Another factor that changes normal workspace requirements is the

dynamic restriction produced by the operator's clothing and personal equipment. For instance, reaching movements are reduced in extent by the wearing of a heavy jacket that prevents complete shoulder movement. Many similar restrictions occur, focused principally at articulation points of the body, and should be accounted for in workspace design. Some movement restrictions are canceled out when, for example, the increased thickness of the uniform pushes the operator ahead on the seat, but if the restriction is severe, it may be necessary to move a control panel closer to the user by some compensating factor. Some clothing restrictions can be reduced by added looseness at bending points, but the side effect of added bulkiness may cancel out the benefit obtained. If the lax material in an elbow joint is loose enough to drag across several controls or catch on protruding objects, it may be more detrimental than the restricted condition that it was intended to cure.

Perhaps the one clothing item that causes more problems in compatibility with workspace than any other is handwear. (See the appendix for sizing information.) Virtually every operator task requires some manual performance, and every glove imposes some interference with the use of the hands. Some gloves give only minimal obstruction, but bulkier items such as dipped impermeable gloves for handling corrosive chemicals and handwear developed to provide cold-weather protection are major offenders. Heavy insulation severely degrades manual performance. Not only does it reduce sensory input from objects touched by the hand, but also the gloves' looseness causes some shifting of the material of the glove when an object is gripped by the hand. Also, since the insulation is larger in overall size than the fingers, the end of the glove bends over before the finger tips can reach the object. All this makes for a clumsy situation and causes significant reductions in operator capability.

At this point it might well be asked why a compromise cannot be made between the maintained warmth of the hand and the amount of insulation used, that is, why the hand cannot be kept cool or chilly and the bulk of insulation reduced. Unfortunately, the performance improvements that one would attempt to achieve with reduced insulation are severely attenuated by cold exposure. At the Quartermaster laboratory, an extensive investigation was conducted of cold-exposure effects in a task requiring manipulative skill [1]. With more than 4,000 trials per subject, the study showed that the hands must remain at a hand-skin temperature of 60°F if adequate performance is to be maintained. Once performance decrements occur, continued cold exposure causes a sharp increase in the rate of failure. Hand-skin temperatures of 55°F produce marked performance deteriorations that hamper operations as much as the insulation would, with the added risk of cold damage to the hands.

Therefore, some device or technique must be developed for maintaining the hands at temperature levels that are comfortable to the man. Comfort is not luxury in this case; when performance is concerned, such temperatures are practical.

What can be done to remedy the variety of factors in handwear that impede manual performance? One possibility is to make the glove an ensemble of subassemblies. This is done in the present standard Army arctic mitten, a knitted wool mitten with a separated index "trigger" finger, which slides into a heavy, lined leather outer shell with a high gauntlet. This item may seem clumsy, but, ~~first of all~~, it is warm. It will protect an immobilized man for hours. Furthermore, it is adequate for rifle carrying and sled pushing, and with some practice a man can do most gross tasks. In short, it functions well as a field-combat protective item for the Arctic, since it serves its primary purpose, protecting the man's hands from the cold. Some finer manipulations can be performed when the outer shell is removed and only the wool insert is worn. However, here there are a number of shortcomings. Probably the worst is associated with rewarming between work periods. First, when the outer shells are removed, the man's hands cool down and so do the shells. Thus, rewarming must be done by putting cold hands back into cold shells. Second, successive rewarming periods are less and less efficient. Eventually, the man's hands and body cool down too much for the hands to be rewarmed at all. This effect is accelerated when the man handles cold-soaked objects as part of his work task. Nevertheless, as has been said, the arctic mitten does fulfill its purpose in many ways. It provides good protection for combat soldiers in the field, it requires no complex, delicate supporting equipment to maintain its operation, and it is a ruggedly built item that will take a lot of punishment.

Where less protection is needed, there is an ensemble consisting of a five-finger leather glove with a five-finger knitted wool insert. This item gives more hand mobility and imposes less obstruction to sensitivity, but it does not give the protection necessary for the Arctic. Therefore, the problem of the cold-weather operation of equipment with heavy handwear is still unsolved.

Wherever heavy handwear must be used in operating conventionally sized controls, such as switches and knobs, task efficiency will be severely affected. A man can be trained to do such tasks with heavy gloves, but the practice time necessary is high, and errors are very likely to occur. For one thing, ganged knobs are very difficult to handle. The setting of one gang frequently causes the glove to rub against other gangs, thereby producing missettings. A knob designed to prevent this in operation with the bare hand may be completely unsuitable for use with gloves. In many cases, toggle switches, and the like, cannot even

be felt through such handwear and have to be snagged by the glove rather than pushed through the glove. Because of this, controls adjacent to the one being operated may be snagged and an error caused. Thus, if controls must be operated by heavy handwear, as is the case with exposed equipment in the cold, the control layout will require considerable space between controls and the use of large-size controls wherever possible. Research has indicated that control knobs smaller than 1 in. in diameter cannot be operated effectively with the Army arctic mitten. Strangely enough, this diameter seems to be the optimum size. No significant performance improvements are obtained by increasing knob diameter to more than 1 in.

Work is also progressing on other methods for protecting the hands without obstructing manual performance and sensitivity. One possibility is that of distributed insulation. There is no necessary reason for making a glove in the age-old tradition of uniform insulation all over the hand. Insulation might be more effective if it were concentrated over the back of the hand, where it would not interfere with performance requirements and could be adjacent to one of the major blood-supply areas in the hand. In a study conducted at the Quartermaster laboratory, there were indications of significantly slower finger cooling with insulation pads on the back of the hand than with equivalent protection in other locations [2]. Further tests are necessary before a conclusion can be reached. However, if such a notion is valid, the performance restrictions of the gloved operator can be much relieved and console design can make much more effective use of the available space.

Study has been begun by the Quartermaster Corps of another possible technique for providing gloves with decreased bulk and improved performance characteristics. This is the by no means new notion of electrical heating. Its main disadvantage is obvious—the need for an accessory power supply. For an infantry soldier burdened by the weight of other gear, the weight allowed for a battery must be small. However, for a man operating exposed equipment, handwear electrically heated by means of an installed generator is another matter. Here the handwear would be operational as long as the equipment worked, and conventional gloves could be issued in case of a power failure. This would be a boon to the hardware designer. A soldier could operate exposed equipment in the cold with little hardship or manual-performance decrement, and the equipment would have to be modified very little, if at all, to fit requirements for optimum manual performance.

Although new approaches are receiving some study, present problems still center around the use of the standard Army arctic-mitten ensemble for operation in the cold. In this situation, as in others where bulky handwear is involved, the separation of control units as much as is

practical and the use of control units of larger sizes can show significant increases in what the gloved operator can accomplish. Since the operator is impeded in everything that he does with his hands, the time allowance for completing tasks should be increased. If countdown procedures are used, where each step in the procedure is completed in order, the time allowed for completion may have to be lengthened to allow for the increased awkwardness of the operator.

Another important problem area in investigating the compatibility of the operator's personal equipment with his workspace is that of headgear. (See the appendix for sizing information.) The principal military requirements for headgear are physical protection from crash and ballistic forces, protection from toxic agents, optimum capability for sight, hearing, and speech, and comfort and fit for the active user. The Quartermaster Corps supply system contains a whole family of headgear items that have been developed to meet specialized requirements of military situations. These headgear items exemplify the compatibility problems that arise when headgear interacts with other equipment in the workspace.

The hard-shell helmets are the principal source of trouble, both because of their size and rigidity and because of problems that occur coincidentally with associated equipment that they contain. The simplest of these helmets is the M-1 infantryman's steel helmet. Its only real complication is that it increases the soldier's overall size. When used with cold-weather clothing with the parka pulled up over it, it causes a considerable increase in the size of the equipped soldier. This must be taken into account in designing other equipment used by the field soldier. For example, the access ways of personnel carriers used in the Arctic may have to be made larger than normal in order to accommodate such troops.

Two other helmets produce more complex compatibility problems. These are the combat-vehicle crewman's helmet for armor personnel and the flying protective helmet for Army flight personnel. Not only are these devices more complex in themselves, but they are used in machine surroundings that present limited working space.

The combat-vehicle crewman's helmet is, first of all, a large item. Its size is necessary to allow space between the outer shell and the head for ventilation and for impact-absorbing material. The space separation is preserved by a sling suspension—a web inside the helmet that fits on the head and supports the helmet shell. In the event of a crash or hit by a projectile, the physical forces are attenuated by the helmet before they reach the skull. The space is also occupied by the earphones of a communications kit that is an integral part of the helmet. These earphones serve in a dual capacity: they act as part of an inter-

communication system, and they form part of a system for noise suppression. The latter is a very important consideration, since noise inside armored vehicles reaches injurious levels, against which man needs some protection. The earphones project up into the helmet proper and extend below its main edge behind earpieces that are part of the helmet shell.

All these characteristics must be borne in mind in the design of the interior of armored vehicles. Since the helmet must be separated from the head to some extent, its size is not likely to be reduced very much in future designs. The problem is made even more difficult by the helmet's rigidity. It does not yield in contact with other objects. This means that the man's head can approach other objects to only a certain limit. The design of aiming and sighting devices should take this into account. If the operator must get his eyes close to an optical sight in order to operate it properly, his efficiency will be much reduced if the helmet bangs up against surrounding surfaces and prevents his eyes from getting into operating position. Eyepieces may have to be extended to reach in under the forward edge of the helmet shell. Also, since the helmet sits up off the top of the head, the requirements for overhead clearance in the workspace should be considered carefully. Any need for the operator to slump in order to operate properly may seriously degrade his performance. If he must slide his body down in order to get his head into position, the relation of the rest of his body to the workspace layout may be seriously disturbed. When cold-weather clothing is added to such a cramped operator, he may be rendered incapable of performing at all.

Most of these problems occur with use of the flying protective helmet as well. It is similar in construction to the combat-vehicle crewman's helmet, except that it fits the head snugly rather than being mounted by a sling suspension. However, the interior impact-absorbing insulation causes the exterior size of the helmet to be considerably larger than the normal bare head. The communications gear is built in, but the headphones do not fit snugly, for the stability of the helmet is independent of the tension of the earpieces against the head. Also, a sun visor is built into the front in such a way that it can be slid down over the eyes or retracted as desired. Otherwise, the helmet's general size is about the same as the armor helmet's, or at least large enough to cause the same space problems. Again, the helmet shell is rigid.

These two headgear items exemplify the problems involved in workspace sizing. When it is necessary to use large helmets whose size cannot be reduced very much, the workspace layout should be reviewed carefully. If the addition of the helmet causes the man to be jammed into the available space, he may be completely unable to perform his

task because he cannot move properly. At best, such space restrictions will cause quick reductions in efficiency, since fatigue effects will be heightened. Therefore, the addition of necessary overhead space to accommodate the operator's personal equipment may pay great dividends in operator performance.

Another important change that occurs when a helmet is placed on the head is a reduction in the visual field available to the wearer. This restriction may be zero or negligible in some cases but can cause a large decrement in operator performance when the restriction is severe. As ballistic and crash protection is increased, the tendency to obstruct vision mounts, affecting the character of the output performance. The armor and the flying helmets that have just been discussed produce such effects to some degree.

Such visual restriction has an important bearing on workspace design. If the operator habitually looks ahead, those displays which require frequent checking should be placed so that the operator can see them within the limits of his restricted visual field while he is wearing the headgear necessary to his task. This consideration is perhaps especially applicable to aircraft, where speed of response is important and becomes more critical as flying speeds are increased. While helmets produce restriction, gas masks and impermeable hoods designed to be used with the helmets usually produce even more. In most cases, the visual restrictions caused by these items cannot be eliminated immediately; so reduced visual capability must be anticipated for operators who are helmeted and are expected to operate masked as well.

Some visual restrictions will probably always be present. It is true that every effort is being made to improve the efficiency of headgear and to increase its protective capacities; however, some physical properties will always act to limit what can be accomplished, because they are basic requirements of the equipment. For example, if crash protection is necessary, the helmet has to overhang the head to some extent; otherwise it loses a great deal of its effectiveness as a head protector. In cases such as this, a compromise must be made between the best possible helmet configuration and the most effective control-display environment in order to permit both factors to interact as efficiently as possible. Some impairments will probably never be eliminated. The most that can be done in these cases is to reduce the impairment to a minimum.

There is still one other factor that has implications for workspace design. This is the need for speech communication and noise suppression, which was mentioned earlier. Not only is loud noise capable of producing hearing loss when it is chronically sustained, but also it interferes with good speech communication. Therefore, since armored vehicles and

aircraft both generate high noise levels, operating personnel need noise protection for reasons of both safety and efficiency. The combat-vehicle crewman's helmet and flying protective helmet do incorporate noise-protection systems, but improvements are still necessary. They suffer from a problem that seems to be common to most helmets of this type. Although they give good overall noise protection, sound frequencies below 1,000 cps are not adequately attenuated, particularly those around 400 cps. This shortcoming seems to be caused by the fact that the helmet shell and the air cavity between the head and the shell resonate at approximately this frequency. Unfortunately for helmet design, a major component of vehicle noise (e.g., engine growl) is a range of frequencies very close to 400 cps. The problem is made even more serious by the fact that such vehicle noise is usually continuous. In a long-term comparison study of the effects of continuous and impulse (short-burst) noise upon hearing thresholds, findings indicated a significantly higher temporary threshold shift for continuous noise [3]. Evidently this type of noise is more disruptive to hearing and may possibly have more serious permanent effects.

Since present helmet designs provide acceptable protection against the higher frequencies, the Quartermaster Corps design efforts are now centered around the provision of better protection for the low-frequency sounds that filter through the helmet. There are several possible methods for achieving this. Their general direction is toward providing more noise resistance through treating the space between the helmet and the head and redesigning the earpieces.

Such efforts will probably not reduce the size of present headgear designs. This fact has important implications for the layout of the workspaces of armored vehicles and aircraft. In providing essential noise protection, it will not be possible to make great reductions in headgear size. Yet the size problem will become more severe as vehicles become larger and more powerful, since higher noise levels will usually be involved. Therefore, unless sound-suppression equipment is incorporated in vehicles to reduce noise levels at the source, helmet designs will have to provide increased noise protection, which almost surely means that they will have to increase in size. Furthermore, there is an upper limit to the noise protection that helmets can provide. Noise increases beyond these limits will have to be reduced by vehicle modification. However, since vehicles are finite bodies and sound insulation would have to occupy space that is already pretty scarce for other equipment, it seems likely that a good bit of the space for extra sound insulation will have to be provided by borrowing from the operator's workspace. This will increase the burden of designing an optimum workplace in which the available space is used to maximum efficiency.

APPENDIX

I. Arctic Uniform

The Quartermaster Research and Engineering Command has published two handbooks containing sizing information relevant to the Army arctic uniform:

Kobrick, J. L.: QM Human Engineering Handbook Series: I. Spatial Dimensions of the 95th Percentile Arctic Soldier, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-39, September, 1956.

———: QM Human Engineering Handbook Series: IV. Dimensions of the Lower Limit of Body Size of the Arctic Soldier, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-51, April, 1957.

The sizing data were obtained by means of a photographic technique that was considered to have general application. Models representative of the upper and lower limits, the 5th and 95th percentiles, of the body-size range of the Army population were dressed in the standard arctic uniform. They were positioned so as to be in the same vertical plane as a size-reference scale graduated in inches and centimeters. Poses were selected in such a way that the maximum number of usable dimensions would be represented in the photographs, and all poses were photographed in front, side, and rear views. The engineer can use these handbooks merely by looking for a picture of the dimension he needs, bridging it with a pair of dividers, and referring the dimension to the size-reference scale shown in that picture. The values are only approximate, with some error due to optical parallax and paper distortion. However, the maximum amount of such error is only about 2 per cent, less than was obtained in direct-check measurements of the subject with an anthropometer, in which the clothing is compressed slightly when contacted by caliper blades. The photographic technique used also took into account the body-movement restriction produced by the uniform. Since the model made his movements while wearing the uniform, output values were restricted when photographed.

Such a photographic technique of size representation is adaptable to other clothing. The method is a timesaver because it avoids the labor of taking and cataloguing multiple measurements. Also, it provides the engineer with considerable latitude of information. He can use any measurement that he sees in the pictures. This leaves to the engineer the decision as to what measurements are of interest to him. If the poses are properly selected, the information should cover changing needs as well.

II. Arctic Handwear

Since significant reductions in the size of arctic handwear cannot be made immediately, the Quartermaster Corps has made available the actual sizes of this handwear, in handbooks of the same type as the arctic-uniform handbooks:

Kobrick, J. L.: QM Human Engineering Handbook Series: II. Dimensions of the Upper Limit of Gloved Hand Size, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-41, December, 1956.

———: QM Human Engineering Handbook Series: III. Dimensions of the Lower Limit of Gloved Hand Size, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-43, February, 1957.

These give the size range of the gloved hand in the Army population for various standard Army handwear ensembles. The photographic method used for the arctic uniform was employed. The hands are shown in frequently used positions for gripping handles and knobs of diameters of ¼ to 2 in. These handbooks give a close estimate of how much space to allow for the use of these handwear systems.

III. Headgear

Recognizing that sizing is a major problem produced by Army helmets, the Quartermaster Corps has published a sizing handbook for these items also:

Kobrick, J. L.: QM Human Engineering Handbook Series: VI. Size Limits of the Head and Neck Area of the Soldier Wearing Army Headgear, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-107, March, 1959.

The photographic method described previously was used again in the same way. The handbook describes the size range of the Army population as it is affected by the major Army headgear items that are likely to cause sizing problems.

The Quartermaster Corps has also tried to help engineers and designers to assess for their own needs the areas that men may be expected to see while wearing present Army equipment. Measurements have been made of the size and shape of the visual fields available to operators with normal vision when they are wearing standard Army headgear items. Included among the items measured were headgear ensembles in which gas masks and protective hoods were combined with the helmets for which they were designed. Measurements were made by means of a large perimeter, with the man's head facing rigidly forward but with the eyes allowed free movement. This technique, which is generally used to measure peripheral vision, was intended to simulate the visual field that would be available to the operator without head movement. The results have been published as part of the Quartermaster Human Engineering Handbook Series:

Kobrick, J. L., and B. Crist: QM Human Engineering Handbook Series: VII. The Size and Shape of the Available Visual Field during the Wearing of Army Headgear, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-133, May, 1960.

The report contains graphs of the shape of the visual field for each headgear item and also lists the tabular values of the measurements. This infor-

mation enables the engineer to tell how much restriction will occur and how it is distributed when the operator is equipped with the headgear used in his task. The data can be applied to any headgear situation and give good guidance to workspace design.

BIBLIOGRAPHY

1. Clark, R. E.: The Limiting Hand Skin Temperature for Unaffected Manual Performance in the Cold, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-147, February, 1961. [Also in *J. Appl. Psychol.*, vol. 45, pp. 193–194, 1961.]

2. Cohen, A.: The Effect of Insulating Palm and Back of Hands on Rate of Finger Cooling, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-153, July, 1961.

3. ———: Temporary Hearing Losses for Protected and Unprotected Ears as a Function of Exposure Time to Continuous and Impulse Noise, *Quartermaster Research Eng. Command Environmental Protection Research Div. Tech. Rept.* EP-151, July, 1961.

Chapter 10

THE DESIGN OF BODY SUPPORT AND RESTRAINT SYSTEMS

*Charles A. Dempsey**

A BODY-SUPPORT SYSTEM IS A MECHANICAL DEVICE THAT SUPPORTS THE BODY in normal and abnormal gravitational environments, whereas a body-restraint system is a device that prevents the body organs as well as the body as a whole from being displaced in the workspace. They can be distinguished by the purpose for which each type of structure is used. The design of these systems requires knowledge of human response to the rigors of both static and dynamic environments. A static environment is defined as a long-term 1-g condition. A dynamic environment is one in which the g forces are continually changing.

This chapter discusses man's physiological and anatomical responses to static environments, his responses to the dynamic environments that are anticipated in space flight, and some of the engineering contributions that have been made to the design of support and restraint systems.

MAN'S RESPONSE TO STATIC ENVIRONMENTS

The prime objective in the design of support and restraint systems for use in static environments is the prevention of sustained localized pressure patterns on the body. Physiological fatigue arising from continuously maintained body positions is due mainly to compression loading of the flesh and the stagnation of body fluids caused by inactivity of the major body segments. In the seated position, physiological fatigue becomes a major problem within a relatively short period of time. The body in the reclining and standing positions is much less susceptible to compression loading and physiological stagnation. In the reclining position, compression loading of the flesh is significantly reduced because the entire body is used to support the individual's weight. For example, the compression loading for a 200-lb man in a reclining position is approximately 0.16 psi, whereas it is 37.0 psi in the seated position. In the stand-

* C. A. Dempsey Associates, Dayton, Ohio.

ing position, compression loading of the flesh is confined to the bottom of the feet. Physiological stagnation occurs when a standing individual remains stationary, but it can be reduced to insignificant proportions by locomotion.

In order to study the localized pressure patterns on the body without exposing human subjects to the dangers of overexposure to X-ray radiation, a mechanism has been designed by Dempsey for the continuous viewing of pressure patterns on any part of the body.[1] The Buttockscope, shown in Figure 10-1, utilizes the technique of edge lighting and the

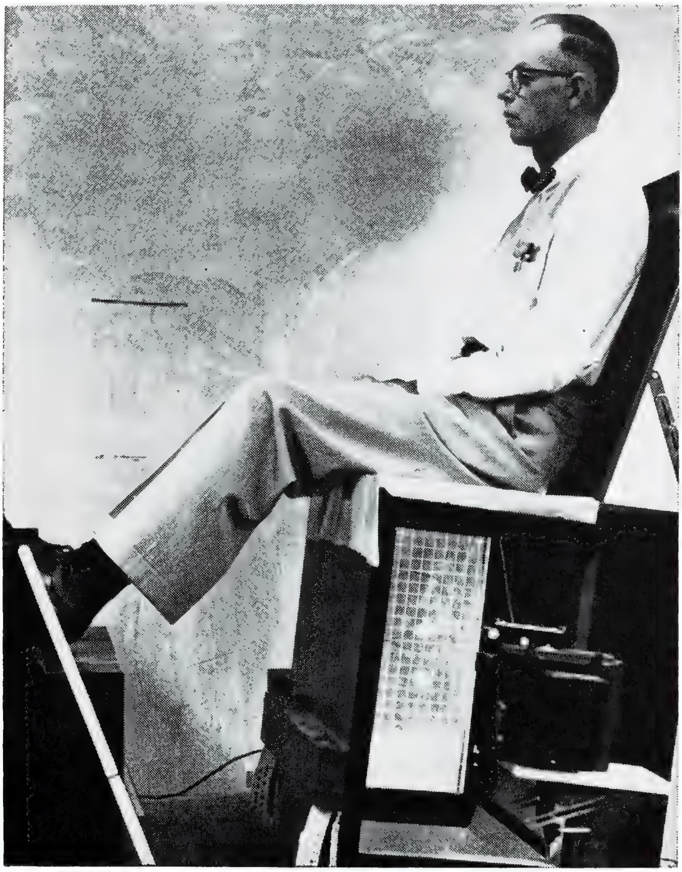


FIG. 10-1 *Subject seated on the Buttockscope.*

changes in light reflection that are caused by compression of a soft material such as "dental-dam" latex sheet material. With a mirror placed immediately below the clear plastic and the latex sheet material, it is

[1] The following information about this device has not been previously published.

possible to measure accurately the area and the distribution pattern of pressure on the flesh and the anatomical support. The seat pan and seat back of the Buttockscope can be adjusted independently, and the distance from the seat reference point to the floor can be changed; thus it is possible to place the man in any of a large number of seat positions.

Experimentation with the Buttockscope has demonstrated that the seated human body supports approximately 75 per cent of the total body weight on 4 in.^{2} of the ischial tuberosities and the overlying flesh, as indicated in Figure 10-2. This load is sufficient to produce compression

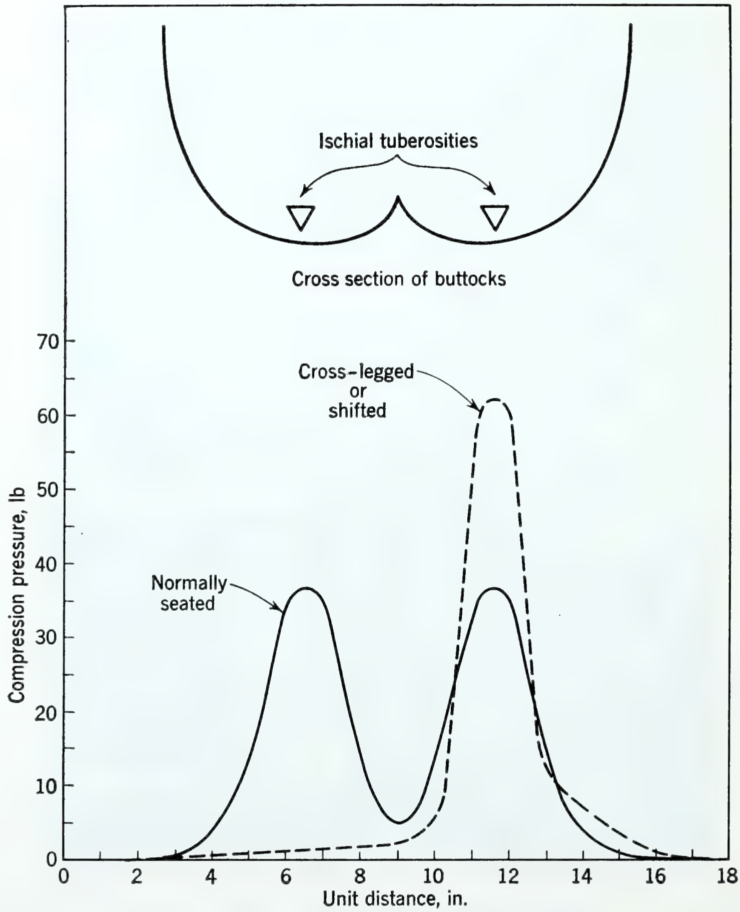


FIG. 10-2 *Pressure-distribution pattern on buttocks during normal seat position and during shifted or cross-legged position.*

fatigue, which varies according to the compression load per unit area and the duration of loading. In physiological terms, compression fatigue is

the reduction of blood circulation through the capillaries, which affects the local nerve endings and results in sensations of ache, numbness, and pain.

Confining human subjects in seats at various angles for periods of 8 hr has demonstrated that ache occurs in the area of the ischial tuberosities after approximately 2 hr, numbness becomes apparent at 3½ hr, and pain appears after 5 hr [18]. The rate of recovery of the tuberosities and the flesh after a 4-hr compression load is slow, since the amount of blood in the capillaries of the entire pressure area has been reduced and must be replenished.

Blood circulation to the capillaries in the buttocks is also reduced through the lack of body-segment movement at regular intervals. If the body is confined in a relatively fixed seating position for more than 4 hr, the physiological functions that control the flow of body fluids slow down and there results a decrease in cardiac output and circulatory exchange [9]. This action coupled with continuous pressure loading on the flesh accelerates the rate of compression fatigue. Although squirming within the fixed seat may delay the slowing down of bodily functions, it can by no means prevent the onset of fatigue.

Fatigue can be delayed, however, by periodic movement of all major body segments, which results in changes in loading conditions and per-

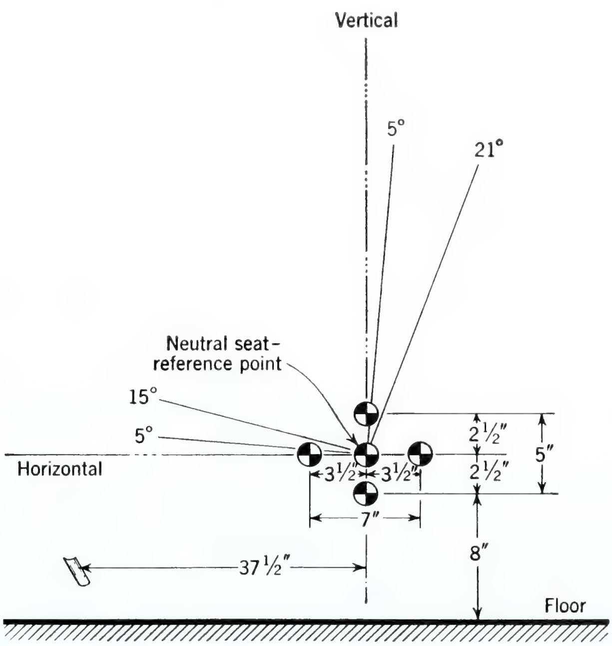


FIG. 10-3 *Seat adjustment for relief of compression fatigue.*

mits muscular expansion and contraction for adaptation to new weight conditions. Extensive research [9, 25] with seats that can be adjusted vertically and horizontally and have independently movable seat backs and seat pans has indicated that the seat-back angle should be adjustable through a range of 5 to 21° aft of vertical and that the seat pan should be independently adjustable through a range of 5 to 15° above the horizontal. In addition, the seat assembly as a unit should be adjustable through a range of 5 in. vertically, with the lowest point of seat reference not less than 8 in. above the floor, as shown in Figure 10-3. With such adjustability it is possible for the occupant to remain seated for extended periods of time without undue slowing down of bodily functions. Generally, men using such a seat will change the position of the seat at least once every 5 min [25]. The type of movement is determined by the type of work required of the individual. In situations requiring alertness, the man will assume an erect posture of 5° back angle, 15° seat-pan angle, and vertical adjustment in the highest position. During periods of minimal activity the posture is 21° back angle, 5° seat-pan angle, and vertical adjustment in the lowest position. The optimum angles for fixed seat back and fixed seat pan have not yet been determined. It is clear, however, that independent adjustment of these two components is necessary.

MAN'S RESPONSE TO DYNAMIC ENVIRONMENTS

The protection of man in the acceleration conditions that he might encounter in space demands the study of human response to such stresses and the evolution of new design concepts for support and restraint. The accelerations encountered during space travel can be divided into three types, vibration, abrupt acceleration, and long-term acceleration. The most dangerous of these is abrupt acceleration, in which the time period from routine conditions to catastrophe is measured in milliseconds and the acceleration magnitude on the periphery of technical measurement.

Various maneuvers of manned space flight, such as soft-surface landings, emergency escape, air deceleration, and hard-surface landings, will produce abrupt accelerations. The potential hazard of explosion is also significantly increased in space vehicles and may form another source of high transient accelerations. Soft-landing impact, ejection, and air deceleration are generally characterized by maximum velocities of 90 ft/sec, with stopping distances of $\frac{1}{100}$ to 20 ft. The relation between velocity and deceleration distance produces accelerations as high as 250 g. Hard-surface landings of space vehicles are usually identified with velocities ranging from 90 to 400 ft/sec. Such landings result in accelerations of 10

to 2,500 g, depending on the type of vehicle attenuation. For explosive forces, the abrupt accelerations extend from 100 g to a maximum approaching infinity. Such accelerations are produced within a time interval of 35 msec from time of ignition to explosion.

The human body may be categorized as a complex of systems that respond to abrupt acceleration forces in accordance with the laws of a viscous-elastic system under the action of a constant unbalanced force or harmonic motion. Although the body may at first appear to act as an integrated whole, it may be segmented into four different parts, each of which seems to respond individually to the force and, in turn, to transmit its response to the other segments. These segments, shown in Figure 10-4,

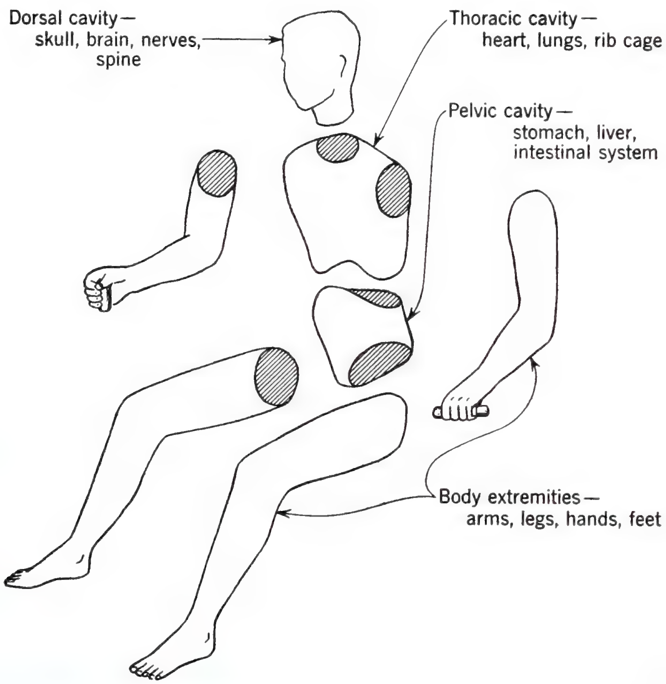


FIG. 10-4 *Four principal segments of the human body.*

are the dorsal cavity (head, neck, and spine), the thoracic cavity (heart, lungs, and rib cage), the pelvic cavity (stomach, liver, and intestinal system), and the body extremities (arms, legs, hands, and feet).

The dorsal cavity is a rigid spherical mass connected to a movable shielded tube that extends along the length of the torso. This entire segment responds to abrupt acceleration forces in a manner similar to that of rigid structural objects and is well protected by structure, liquids, and

TABLE 10-1 *Test Values of Human Tolerance to Accelerations and Type of Restraint Devices Utilized*

Body units	Direction of acceleration	Physiological description of acceleration	g	Duration, sec	Onset, g/sec	Restraint system
Head	Upward	$+G_z$	16	0.04	......	Face curtain [26]
	Forward	$+G_x$	22	....	600	None [24]
			25	1	600	Helmet and straps [22]
			45	0.044	500	None, against chest [22]
	Downward	$-G_z$				
	Backward	$-G_x$	83	0.04	4,000	Rigid headrest [2]
			12	5	......	Helmet and straps [5]
			10	23	......	Helmet and straps [5]
Torso*	Upward	$+G_z$	28	....	......	Upper-torso harness [17]
			40	....	......	Torso restraint F-106A [20]
	Forward	$+G_x$	5.5	13	......	Lap belt, shoulder harness, chest strap, helmet tie-down, crotch strap [19]
			25	1	600	Shoulder harness, lap belt, inverted-V crotch [24]
			18	0.065	......	Abdominal vest [3]
			45	0.044	500	Stapp's restraint gear (Group H) [22]
	Downward	$-G_z$				
	Backward	$-G_x$	83	0.04	4,000	Group H harness [2]
Arm	Upward	$+G_z$	16	0.04	......	Face curtain [26]
	Forward	$+G_x$	25	1	600	Elbows tied behind back, wrists tied at knees [24]
	Downward	$-G_z$				
	Backward	$-G_x$				
Leg	Upward	$+G_z$				
	Forward	$+G_x$	35	0.1	1,156	Toe straps, canted footrests [22]
			25	1	600	Knees and ankles tied together, feet tied to footrests [24]

Leg (cont.)....	Downward	$-G_z$	11	....	......	Belt over thighs [15]
	Backward	$-G_x$	35	0.1	1,156	Toe straps, canted footrests [22]
Entire body...	Upward	$+G_z$	16	0.04	......	Lap belt, shoulder straps, face curtain [26]
			28.5	0.020–0.033	......	Safety belt [14]
			5	240	0.07	Anti-g suit [20]
			16	5	......	Water immersion [7]
	Forward	$+G_x$	8	1	......	Bernardini's suit (with vibration at 0.7 cps) [21]
			12	4	......	Bernardini's suit [5]
			10	120	......	Supine—net seat [1]
			45	0.044	500	Stapp's restraint gear (Group H) [22]
			55	....	......	Safety belt and yielding structure [12]
			31	5	......	Water immersion [7]
	Downward	$-G_z$	11	....	......	Shoulder strap, belt over thighs, lap belt [15]
	Backward	$-G_x$	16.5	....	......	Supine—net seat [1]
			10	120	......	Prone—net [1]
			14	120	......	Supine, forward tilt, water immersion [4]
			35	0.1	1,156	Lap belt, leg restraint, headrest, chest strap [22]
			31	....	......	Water immersion, supine [26]
			25	....	......	NASA supine couch [7]
			83	0.04	4,000	Group H restraint [2]
			12	5	......	Helmet and straps, lap belt, spine forward 25°, rigid seat, legs sharply flexed [5]
			10	23	......	Helmet and straps, lap belt, spine forward 25°, rigid seat, legs sharply flexed [5]
			16.5	170	......	Contoured net seat, spine forward 12°, legs at 90° [6]
	Upward	$+G_z$				
	Forward	$+G_x$	162	....	22,600	None [13]
			209	0.015	27,800	None [8]
			106	0.024	8,333	None [8]
			33	0.061	1,100	None [8]
			28.5	0.073	780	None [8]

TABLE 10-1 *Test Values of Human Tolerance to Accelerations and Type of Restraint Devices Utilized (Continued)*

Body units	Direction of acceleration	Physiological description of acceleration	g	Duration, sec	Onset, g/sec	Restraint system
Entire body (cont.)	Downward	$-G_z$				
	Backward	$-G_x$	183	0.013	28,000	None [8]
			140	0.012	23,300	None [8]
			123	0.024	10,250	None [8]
			95	0.031	6,100	None [8]

* See also Entire body.

SOURCE: S. Shelton, Personnel Restraint Devices for Advanced Flight Vehicles, *USAF WADD Tech. Rept.* 60-301, pt. 1, June, 1960.

internal flexible attachments. The brain is protected within the structure through multiple flexible attachments and floats within a liquid that increases the natural protection. If the rigid structural processes of the skull and spine are not violated by crushing or puncture, the major threat to the dorsal cavity is the rapid movement of the spine to its angular limit. Such a movement in the neck area, immediately below the base of the skull, has produced concussions in experimental animals by the sudden stretching of the spinal cord [16]. However, damage to the dorsal cavity will not normally occur if the natural limits of movement are maintained and the organs are prevented from reaching a hyperextensive condition.

The thoracic cavity is an entirely different type of structural and physiological system. The bony processes of the rib cage provide some natural protection to the internal organs even though the ribs are flexible and move with the simple act of inhalation and exhalation. The ribs can be expected to carry a limited quantity of force, but the unit-area loading must be small and spread over the entire surface. Fracture of the rib cage inwardly can produce internal organ damage and in some cases may contribute, through puncture, to the failure of the heart or lungs. The thoracic cavity represents a mass within a mass, wherein the heart is an organ that can move like a pendulum within the rib cage. This movement is restricted in the spinal and sternum directions by the rib cage and the spinal column and in the head and foot directions by the diaphragm and upper limit of the chest, but it has been shown in animal studies [16] that in the right and left lateral directions the heart may be displaced to the full limit of the thoracic cavity by forces as low as 10 g. Since it is not possible to restrain the heart directly, damage to this and other organs within the thoracic cavity is always possible.

The pelvic cavity contains no rigid structure, and therefore it functions as a viscous-elastic system. It is able to extend to relatively large distances in six directions (fore and aft, left and right, up and down) and to return to its original position without injury. Because of the pelvic cavity's composition and ease of movement, it is difficult to apply a restraining force to this body segment. All organs within it are soft and displace over one another within the fluid medium. The different organs are joined together by flexible fibrous ligaments, which can sustain injury caused when the differential displacement between organs exceeds ligament length. In terms of empty-organ weight, the liver represents the greatest mass within the pelvic cavity. However, the weight of digested food within the intestinal system usually exceeds that of the liver and therefore determines the response of the other organs. Because individual organs within the pelvic cavity cannot be directly prevented from moving independently of each other, differential-displacement injury is a com-

mon occurrence in impact accelerations. The injury potential can be reduced, however, by maintaining the volumetric form of the pelvic cavity through the application of a low unit-area pressure over the entire body segment.

The body extremities, the arms, legs, hands, and feet, do not contain vital physiological functions, and their primary limitation is the sphere of movement that they cannot exceed without injury. Published anthropometric data have defined the limits of extremity movement [11, 23].

Table 10-1 outlines some test data on human tolerance to both sustained and impact acceleration. See Chapter 4 for further details on the body effects of acceleration.

THE DESIGN OF MECHANICAL AIDS

The design of body-support and -restraint systems for both static and dynamic environments requires an evaluation of the available materials. High-strength rigid materials that satisfy structural requirements are not satisfactory for human habitability or the relief of fatigue. On the other hand, the soft, pliable materials capable of reducing fatigue cannot resist high-stress applications. Therefore, optimum design depends upon the interaction of soft and rigid materials to meet the seemingly incompatible requirements of man and his environment. Table 10-2 reviews the materials that have been used for support and restraint systems.

One method of reducing compression fatigue is the contouring of materials in such a manner that they completely support the entire body surface. Contouring with rigid structural materials involves the difficult problem of adequate anthropometric fit. The solution to this problem seems to be the integration of soft and pliable foams with rigid contour materials or the adjustment of the contours for each individual by the use of stretchable materials, such as nylon Raschel net.

Another experimental approach to the reduction of compression fatigue is the Dynamic Cushion, developed by Dempsey et al. [9]. It was reasoned that, since it is unit loading of the flesh and tuberosities that produces compression fatigue, such fatigue might be prevented if the compression load could be periodically shifted from the tuberosities to other portions of the buttocks. It was realized, however, that a shift in compression loading to other fleshy areas cannot be imposed for protracted periods, for the tuberosities are the primary weight-carrying structures in the buttocks. The device constructed for shifting the compression load was a rubber bladder with holes for the tuberosities. When the bladder is deflated, the occupant is seated on the tuberosities in the normal compression-loading position. When the bladder is inflated, the

TABLE 10-2 *Material Requirements for Seat Design*

Material	Physiological requirements						Mechanical requirements					
	Soft	Pliable	Stretch-able	Porous	Dampen-ing	Formed	Rigid	High strength	Formed	Weight		Dampen-ing
										Low	High	
Steel	...	...	...	...	...	X	X	X	X	...	X	
Aluminum	...	...	...	...	...	X	X	X	X	X		
Fiberglas	...	...	...	...	...	X	X	X	X	X		
Wood	...	...	...	...	X	X	X	X	X	X	...	X
Wire mesh	...	X	...	X	X	X	X	X	X	X	...	X
Nylon net	X	X	X	X	...	X	...	X	X	X		
Sand	X	X	...	X	X	X	X	X	X	...	X	X
Liquids	X	X	...	...	X	X	...	...	X	...	X	X
Inflatable structures	X	X	...	...	X	X	X	X	X	X	...	X
Rubber foam	X	X	X	X	...	X	...	...	X	X		
Fabrics	X	X	X	X	...	...	X	X	...	X		
Plastic spheres	X	X	...	...	X	X	X	...	X	X	...	X

body is lifted approximately ¼ in. and the compression loading is removed from the tuberosities and transferred to the surrounding flesh in the buttocks. Experiments with the Dynamic Cushion established that a square-wave cycle of 20-sec duration would adequately shift the compression loading on the buttocks and prevent compression fatigue, as illustrated in Figure 10-5. The combination of the Dynamic Cushion and

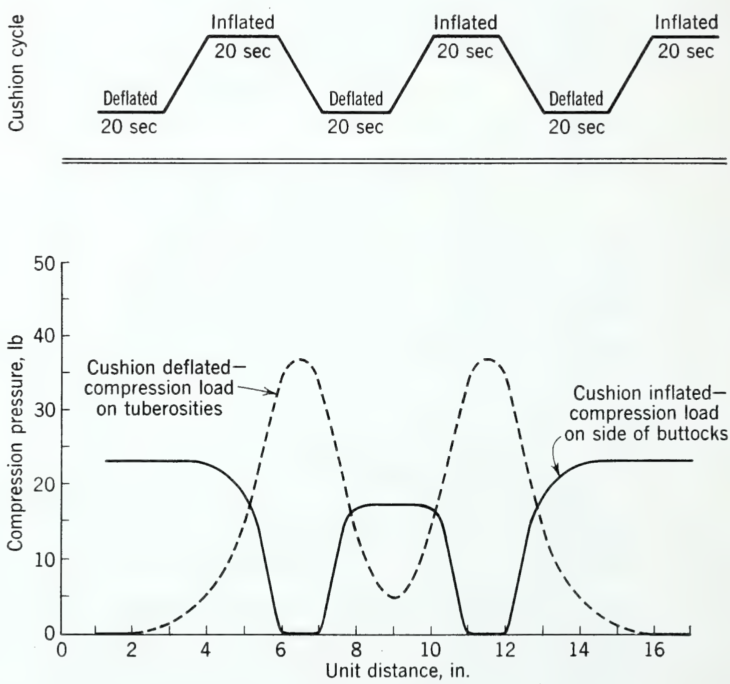


FIG. 10-5 *Functioning of the Dynamic Cushion.*

the tiltable seat has been successfully tested under flight conditions for 80 consecutive hours during which the pilot never left his seat [9].

Seats constructed of Raschel net have supported human subjects for 45 consecutive hours without causing physiological fatigue [10]. In addition, such seats, one of which is shown in Figure 10-6, have safely supported and restrained centrifuge subjects with continuous loadings of 16 g [6]. Figure 10-7 shows a typical piece of nylon Raschel net; the inset shows its detailed structure. Nylon net has high-strength characteristics when the lines of force are properly oriented to the weave of the net. One drawback of nylon net seems to be its inability to control rebound actions under dynamic environmental conditions.

FIG. 10-6 *Supine net seat.*

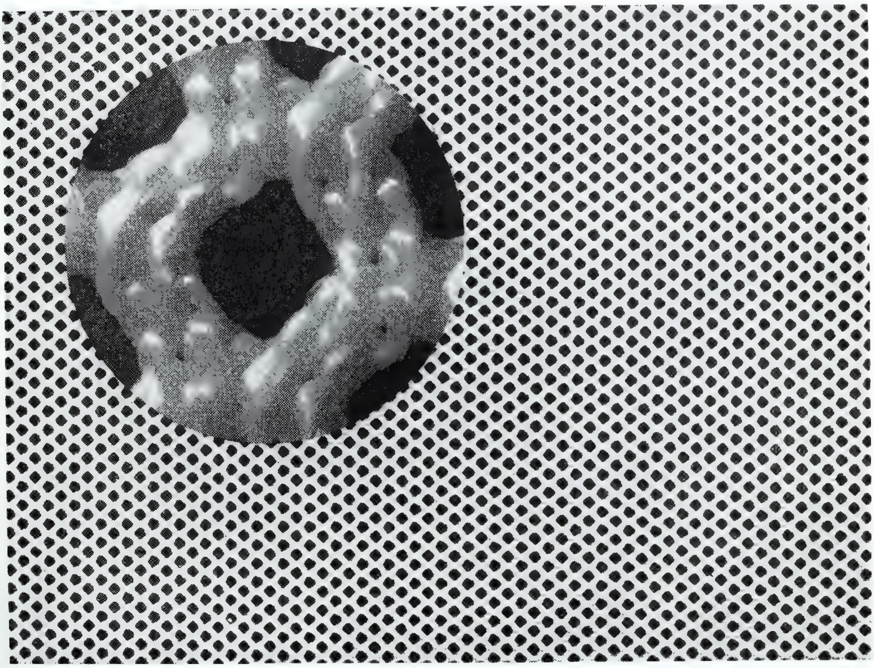


FIG. 10-7 *Raschel net.*

A number of advances have been made in the design of nylon-net seats. One new experimental design of a nylon-net seat has been developed by Bennett and tested by Bennett and Carter while at the Bio-Mechanics Laboratory of Tufts University. This seat, details of which are shown in Figure 10-8, weighs 1½ lb installed. It was designed for situations in which long-duration support, minimum workspace, rapid entrance and exit, and minimum weight might be important considerations. It has been tested without causing physiological fatigue or subjective discomfort over 48 hr of extended seating. The experimental seat design involves a number of innovations developed to test a variety of theoretical questions. Note in Figure 10-8*a* that the seat is hooked to an experimental frame that includes adjustable armrests (shown in Figure 10-8*b*) separate from the seat proper and also a 1-in.-thick foam-rubber floor to supply all leg support through the feet. The seat proper carries body weight exclusively through the back and buttocks, without leg support or pressure on the popliteal area.

Figure 10-8*c* looks down on the seat-pan triangle. Note that the front suspension hook is not through the center grommet but through the two grommets 1 in. either side of center. This distributes the effective suspension across the front 2 in. of the seat-pan front and tends to contour the seat pan to generate force lines across the individual buttocks, relieving the center line of upward pressure and thus decreasing anal compression and discomfort. Figure 10-8*d*, which gives a front view of the back triangle, shows the experimental hook used for height adjustment and for quick engage and disengage for entrance and exit of the seat as part of the operator's clothing. Figure 10-8*e*, which looks down on the left side support of the back and seat-pan triangles, shows the lacing of the first 12 in. of the back and pan grommets. This lacing tends to control further the contour of the pan, running the pressure across the buttocks to avoid anal pressure. Force on the back is mainly down the spine and is partially transmitted through the grommet lacing to help contour and flex the seat pan. Figure 10-8*f* shows the excess back material used to hook the seat to the work suit to make it a part of the work clothing. There is a similar excess at the seat pan. Note that the seat can be attached between the man's back and any required back pack or between his buttocks and any required seat pack, thus permitting him to sit with his pack still attached but outside the seat itself.

Support and restraint systems for dynamic conditions may be either partial or whole-body. Partial-body restraint, in general, consists of a series of straps or small mechanical pads that cover less than 40 per cent of the total body surface and are used to restrain the major body segments and the total body mass. A whole-body restraint encloses the entire body mass in such a manner that differential displacement is impossible and

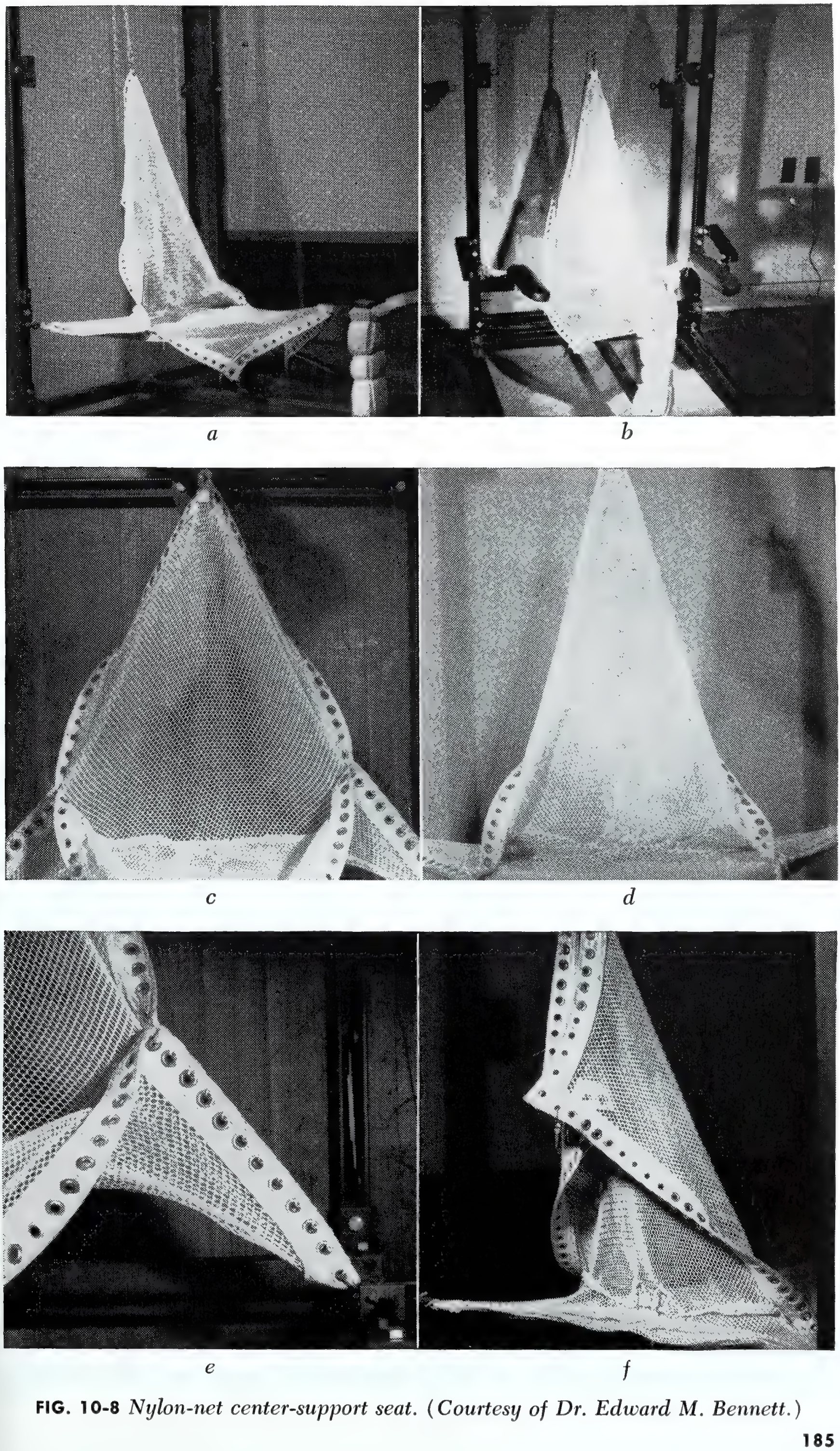


FIG. 10-8 *Nylon-net center-support seat. (Courtesy of Dr. Edward M. Bennett.)*

the occupant is completely rigid in the direction of the imposed dynamic load. Past research has demonstrated that partial-body support and restraint is an effective and economical means of human protection under dynamic conditions that do not exceed 35 g with rates of onset of 1,500 g/sec. In future flight operations, however, the dynamic environment may often exceed these limitations in emergency conditions and therefore may raise the need for whole-body support and restraint.

In order to provide protection for astronauts during abrupt acceleration conditions, equipment must be designed about the following functions: First, an automatic sensing and implementation system is necessary for simultaneously monitoring the vehicle and life-space mechanical functions and detecting the conditions that require the use of protective equipment or emergency-escape devices. Second, an automatic positioning device is required for moving the man and his extremities to the completely restrained body position prior to the initiation of emergency escape or the impacting of a vehicle during a soft or hard landing. Third, the whole-body restraint system must provide a complete and rigid constraint for the man when abrupt accelerations exceed 15 g in any direction. Provision must also be made for the automatic relaxation of the restraint system when the accelerative force has been removed. Fourth, the whole-body restraint should be capable of repeated usage throughout the entire mission. Fifth, the unit loading of the force on the body should be uniformly distributed over the entire body surface and should be compatible with the different physiological responses of the four major body segments. Sixth, the whole-body restraint should not interfere with the mobility of the man during unrestrained flight.

In the pursuit of these objectives, the design of whole-body-support and -restraint mechanisms must protect the man against abrupt accelerations of up to 60 g for time intervals of 50 msec. One current Air Force Aeronautical Systems Division (ASD) research program, using a molded Fiberglas couch similar to the Project Mercury foam-plastic couch (Figure 10-9), is investigating the supine body position, with the acceleration force being imposed in the sternum ($-G_x$) direction. A minimum-configuration lap belt, a shoulder harness, and a headband hold the subject securely in the fully contoured seat and prevent adverse rebound effects. Over 100 experiments using human subjects have been conducted with up to 60 g of abrupt acceleration. It has been demonstrated that with this minimal type of whole-body restraint these accelerations are below the voluntary tolerance of man and do not produce a shock syndrome. A comparison between the physiological response factors and the acceleration-time histories in these experiments indicates that the total experience is well within the latent period of body response. In another ASD program, an entirely new whole-body restraint is being

designed for investigation with all acceleration directions. Human experiments using acceleration levels of 15 g in the foot ($-G_z$) direction, 30 g in the head ($+G_z$) direction, and 60 g in the spine ($+G_x$), sternum ($-G_x$), and right ($-G_y$) and left lateral ($+G_y$) directions are presently being formulated on the basis of knowledge obtained in the experiments with the fully contoured seat. So that the body will be able to withstand such forces without injury, this whole-body restraint will automatically position the man prior to the acceleration and will completely constrain all body segments.

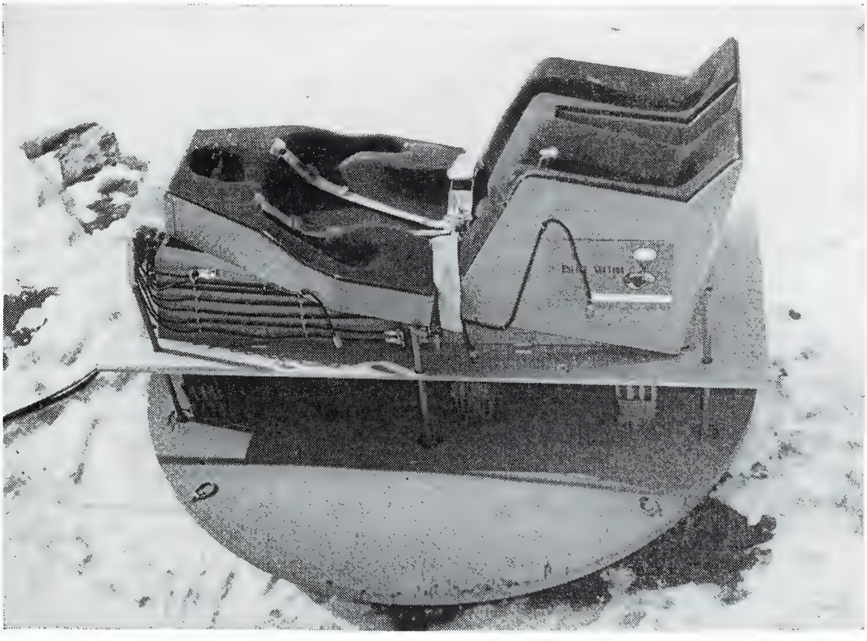


FIG. 10-9 *Project Mercury foam-plastic contoured couch.*

Results obtained in the first set of experiments and preliminary tests in the second set suggest that man can be exposed to progressively higher magnitudes of abrupt acceleration above the 60-g level without major injury if whole-body restraint is used and the total time of exposure to the force is within the latent response period of the body. The magnitudes that can be tolerated are on the lower limit of explosive acceleration forces to be encountered in emergency conditions of manned space flight, but present research suggests that the tolerance limit can be extended much further through the development of protective devices.

BIBLIOGRAPHY

1. Ballinger, E. R., and C. A. Dempsey: The Effects of Prolonged Acceleration on the Human Body in the Prone and Supine Positions, *USAF WADC Tech. Rept.* 52-250, July, 1952.

2. Beeding, Eli L., Jr., and Rufus R. Hessberg, Jr.: Daisy Track Tests 271–337; Run: 4 February–19 May 1958, *USAF Missile Developm. Center Aeromed. Field Lab. Test Rept.* 8, November, 1958.

3. Bierman, H. R., Russell M. Weeder, and Harper K. Hellems: The Principles of Protection of the Human Body as Applied in a Restraining Harness for Pilots, *Naval Med. Research Inst. Rept.* 6, Project X-630, May, 1946.

4. Bondurant, S., et al.: Effects of Water Immersion on Human Tolerance to Forward and Backward Acceleration. *USAF WADC Tech. Rept.* 58-290, July, 1958.

5. Clarke, Neville P., and Stuart Bondurant: Human Tolerance to Prolonged Forward and Backward Acceleration, *USAF WADC Tech. Rept.* 58-267, July, 1958.

6. ——— et al.: A Preliminary Report of the Human Response to Rearward Facing Re-entry Accelerations, *USAF WADC Tech. Rept.* 59-109, July, 1959.

7. Collins, C. C., R. S. Crosbie, and R. F. Gray: Pilot Performance and Tolerance Study of Orbital Reentry Acceleration, *USN Air Developm. Center, AMAL-Letter Rept.*, Project TED-ADC AE-1412, Sept. 19, 1958.

8. DeHaven, Hugh: Mechanical Analysis of Survival in Falls from Heights of Fifty to One Hundred and Fifty Feet, *War Med.*, vol. 2, pp. 586–596, July, 1942.

9. Dempsey, C. A., et al.: The Human Factors of Long Range Flight, *J. Aviation Med.*, vol. 27, pp. 18–22, February, 1956.

10. Duddy, John H., and C. A. Dempsey: Light-weight Seating: Design Research and Development of a Net Seat for Project MANHIGH, One of a Series of Studies Pertaining to Crew Compartment Habitability for Extended Missions, *USAF WADC Tech. Rept.* 58-307, December, 1958.

11. Hansen, R., and D. Carnog: Annotated Bibliography of Applied Anthropology in Human Engineering, *USAF WADC Tech. Rept.* 56-30, May, 1958.

12. Hasbrook, A. Howard: Crash Survival Study of a Survivable Crash Involving a Piper PA11, Aviation Crash Injury Research of Cornell University, New York, May, 1954.

13. ———: Human Impact Survival at 162 g, *Aviation Crash Injury Research Rept.* AV-CIR-58-0-101, March, 1959.

14. ———, John T. Pairn, and H. R. Guggenheimer, Jr.: Av-Air Crash Survival Study; U.S. Army Bell H-13 Accident at Monmouth County Airport, Belmar, N.J., June 21, 1956, *Aviation Crash Injury Research Cornell Univ. Rept.* AV-CIR-5-CSS/H-81, February, 1957.

15. Hecht, Kenneth F., Edward G. Sperry, and Francis J. Beaupre: Downward Crew Ejection Seat Tests from the B-47 Airplane, *USAF WADC Tech. Rept.* 53-433, November, 1953.

16. Hollister, N. R., et al.: Biophysics of Concussion, *USAF WADC Tech. Rept.* 58-193, September, 1958.

17. Kalogeris, J. G.: Pilot Emergency Escape Upper Torso Harness Support, Development Test of Model 106A, *Convair Rept.* 9999, Aug. 28, 1956.

18. Michael, K.: Etiology of Decubitus Ulcers, *Arch. Phys. Med.*, vol. 42, pp. 19–29, January, 1961.

19. Miller, C. O.: Evaluation of Transverse Acceleration (Rear to Front) Utilizing Conventional and Special Restraint Gear, *Chance Vought Aircraft Co. Rept.* 10816, Feb. 20, 1957.

20. Miller, Hugh, et al.; The Duration of Tolerance to Positive Acceleration, *USAF WADC Tech. Rept.* 58-635, November, 1958.

21. Poppen, J. R., and J. H. Modlin: F106A Torso Restraint Systems, Results of Preliminary Study and Test Evaluation, *Radioplane Corp. Rept.* AMM-66, Apr. 8, 1957.

22. Stapp, John Paul: Human Exposures to Linear Deceleration; Part 2. The Forward-facing Position and the Development of a Crash Harness, *USAF Tech. Rept.* 5915, pt. 2, December, 1951.

23. Stoddard, T. L., and W. K. Carter: An Optimum Manual Workspace with a Seated Operator, paper presented at American Psychological Association Convention, Washington, September, 1958.

24. Tobak, M., and H. J. Allen: Dynamic Stability of Vehicles Transversing, Ascending and Descending Paths through the Atmosphere, *NACA Tech. Note* 4275, July, 1958.

25. Van Wart, F., and C. Dempsey: Project Long Gone: A World's Record for Jet Flight, paper presented at American Astronautical Society Convention, New York, January, 1960.

26. Webb, M. J., and R. F. Gray: Protection against Acceleration by Water Immersion. *Am. Rocket Soc. Paper* 805-59, 1959.

Chapter 11

OPERATOR MOBILITY IN PRESSURE SUITS

C. Lawrence Bommarito[*]

AT SEA-LEVEL PRESSURE OF 760 MM HG (14.7 PSI) THE EARTH'S ATMOSPHERIC oxygen is 21 per cent of the total atmosphere, accounting for a partial pressure of 158 mm Hg (3.07 psi). The composition of the atmosphere remains constant to altitudes of as much as 50 miles, but the density is so greatly diminished that life is impossible without the protection of body pressurization to at least 128 mm Hg (2.5 psi) and an oxygen supply at or near the sea-level partial pressure. Currently operational pressure suits provide body pressurization to 35,000 ft or nearly 180 mm Hg (3.5 psi) and 100 per cent oxygen for breathing so that the respiratory supply is nearly the same as at sea level. The body pressurization is required to prevent the vaporization or "boiling" of the blood that occurs at 63,000 ft (47 mm Hg or 0.91 psi). Respiratory assistance above 40,000 ft (140.7 mm Hg or 2.72 psi) requires that oxygen be supplied under pressure in order that transfer of oxygen through the lung tissue to the blood can occur and provide sufficient blood oxygenation.

Normally the pressurization of the cabin will provide a safe, comfortable environment for the space traveler. The pressure suit and its independent oxygen supply can sustain a man in the event of failure of the cabin system. In such an emergency, the crewman of a space vehicle requires as much freedom of movement as possible. However, the pressurization of his protective suit limits his movements considerably. Therefore, in the design of an efficient vehicle, measurements are required of the angles to which a man in a pressurized suit can rotate, flex, and ex-

* The Boeing Company, Seattle, Wash.

The author wishes to acknowledge the assistance provided by the following persons, without whose help the completion of the work would have been impossible: Capt. Melvin Snowden and Lieut. Cmdr. Walter Goldenrath of the Pressure Suit Training Unit, Naval Air Station, San Diego, Calif., who supplied three of the pressure suits used in the study; Elliott Prindle, who arranged for the loan of the additional suit; Wayne Springer, D. P. Turner, K. Pitz, and J. D. McClure, who assisted in the laboratory work.

tend his head, torso, arms, hands, legs, and feet and of the forces that he must exert to achieve these movements.

TEST PROCEDURE

An investigation was made of gross body motions and attenuation of these motions by full-pressure suits. The motions investigated were standard ones, whose reliability and intercorrelations have been established [1]: standing flexion; head movement, ventral; head movement, ventral to dorsal; head movement, lateral; head rotation; upper leg, extension backward. The six body flexions were determined for four subjects in three full-pressure suits and for three subjects in a fourth suit. The tests previously established by Dusek and Teichner were eight in number. Two of these were considered inappropriate for the present study and were therefore omitted.

Suit Pressurization

Each pressure suit was supplied with aviator's breathing oxygen from two independently operated cylinders. One tank supplied 60 psi oxygen to the helmet regulator; the second was used to inflate the suit. The supply from the second tank was routed through a ventilation port in the left side of the suit. The supply of oxygen to the helmet was released about the subject's face "as demanded." A rubber seal ringed the face to prevent gases in the suit from entering the area adjacent to the subject's nose and mouth. When breathing oxygen was required, the helmet passed a fresh supply into the breathing area. Exhausted gases passed through a one-way valve into the suit. Once the suit was fully pressurized, the exhaled gases were of sufficient quantity to maintain the suit pressure.

Suit pressure was indicated by a mercury manometer and was maintained between 6 and 6.5 in. Hg (2.95 to 3.20 psig). Pressurization was controlled manually by regulating the release of excess gas through a vent port in the right side of each suit. This was considered to be the most practical laboratory method, but it did permit some variation in suit pressure during flexion maneuvers, especially during head-movements tests. Changes in body position momentarily caused suit-pressure fluctuations of up to ± 0.2 psi. It was found that instructing the subject to change positions slowly could minimize the pressure fluctuation. The face seal was responsible for part of the limitation of head movements. Tilting and rotation of the head occasionally caused leakage from the breathing compartment into the suit. Once the face seal was "broken," the helmet

regulator opened and allowed a continuous flow through the face mask into the suit.

The intercommunication microphone and earphone receivers were connected to a headset used by the observers. Thus, constant communication with the subject was maintained.

Suit and Helmet Adjustment

The helmet required adjustment to each individual in order to provide an adequate seal at the face. It was extremely difficult to maintain proper sealing whenever head movements were attempted. In practically all instances of extreme head movements (ventral, dorsal, lateral, or rotational) the seal allowed passage of gas from the breathing compartment into the suit. During standing-flexion tests the face seal was easier

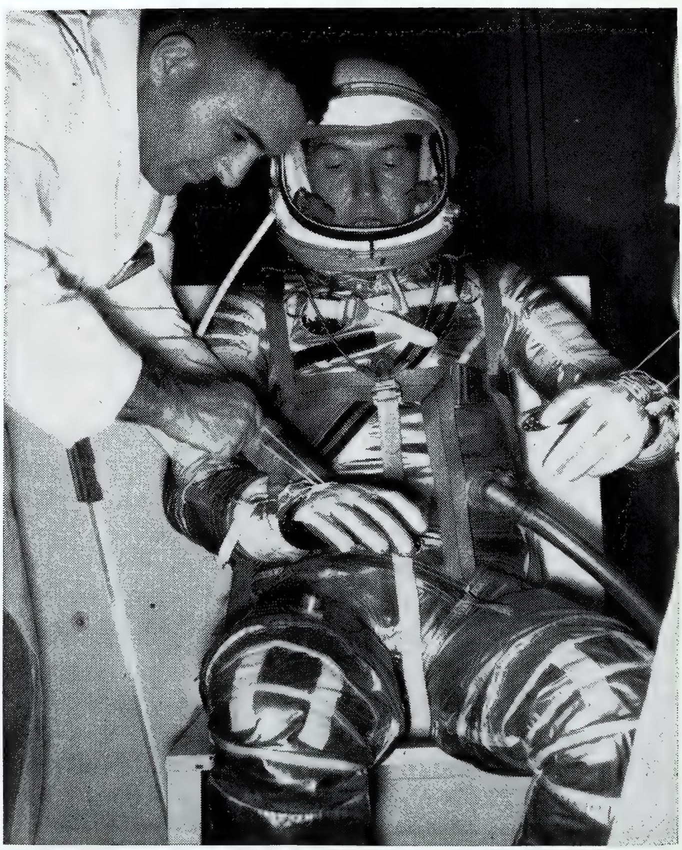


FIG. 11-1 *Detail of helmet tie-down system. Other straps and lacings provide configuration control.*

to maintain, but in some instances leakage did occur. Tightness of the seal during standing flexion depended not only upon proper helmet fitting but also upon the adjustment of helmet tie-down straps. The helmet tie-down system used is shown in Figure 11-1. Minimal adjustments were necessary to fit the suit torso to individual subjects, since the subjects selected were of similar physical dimensions. The subjects' mean height

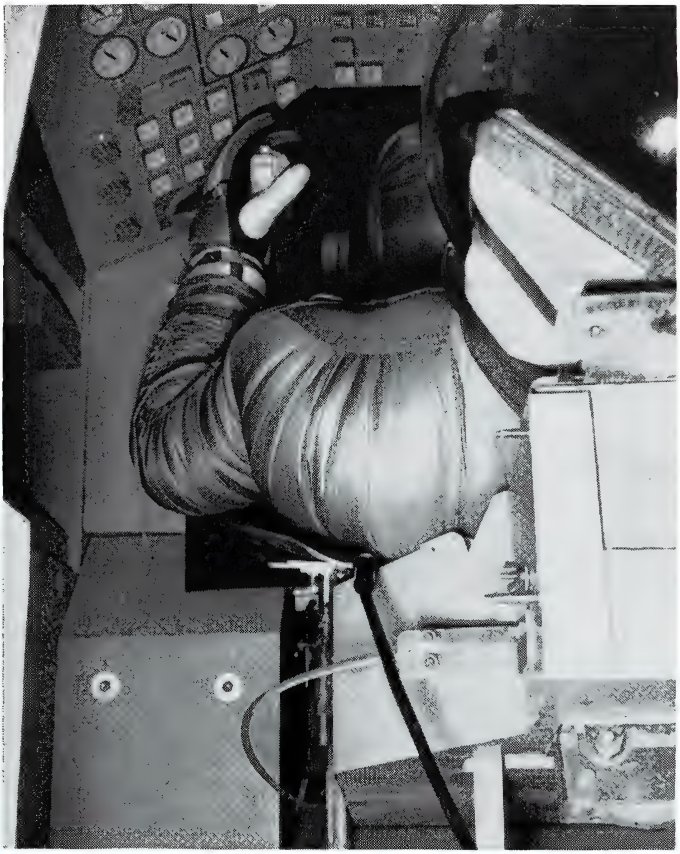


FIG. 11-2 *Detail of shoulder and arm configuration during pressurization.*

was 67.7 in., mean weight 170.7 lb, and mean age 27.5 years. Figure 11-2 shows the shoulder and arm configuration in the pressurized suit.

Body Flexion

Gross body motions were determined for the four subjects in three conditions: (1) loose-fitting coveralls for control; (2) pressure suits at 0 psig; (3) pressure suits at 3.0 psig.

The six tests utilized were as follows:

1. *Standing Flexion.* The subject stood with his toes on the edge of a box and reached downward as far as possible without bending his knees. The position was held for 5 sec while the experimenter recorded the measurement.

2. *Head Movement, Ventral.* A goniometer for measuring angular motion was attached to a lateral surface of the subject's head and was set at zero, with the subject seated and his head in a normal upright position, upper jaw parallel to the floor. The angular extent of tilt of the head was then recorded when the subject moved his neck in ventral flexion as far as possible without moving the chest or shoulders.

3. *Head Movement, Ventral to Dorsal.* A goniometer was attached to the lateral surface of the subject's head and was set at zero when the subject was seated and his neck in ventral flexion. The subject then extended his head as far back as possible without moving his shoulders. The angular extent of head movement was recorded.

4. *Head Movement, Lateral.* A goniometer was attached to the dorsal surface of the subject's head and was set at zero when the subject was seated and had moved his head as far to his right as possible without having moved his shoulders. The subject then moved his head to the left as far as possible. The extent of the angular movement from right to left was recorded.

5. *Head Rotation.* The subject stood and bent his back forward at the waist while holding the seat of a chair so that his back and neck were parallel to the floor. A goniometer was attached to the cranial surface of the subject's head and was set at zero when the subject had rotated his head to the right as far as possible. He then turned his head to the opposite side as far as possible, and the extent of rotation was measured.

6. *Upper Leg, Extension Backward.* A goniometer was attached to the left upper leg on the lateral surface proximal to the knee. The subject faced and touched a wall, with his left hip and leg extending slightly beyond into an open doorway. The goniometer was set at zero, with the subject pressing his left hip against the edge of the wall and with his left leg extending backward and upward while the knee was kept straight and the left hip in contact with the edge of the wall. The position was held while a reading was made.

The "control" tests were those performed while the subject was attired in loose-fitting coveralls. In all cases the control tests were performed prior to the subject's donning any of the three pressure suits. The test run in the unpressurized state of each suit was obtained before pressurizing the subject in that suit.

During control tests and unpressurized runs the testing proceeded in the order of the tests listed previously. During pressurized runs, how-

ever, it was found that control of the suit pressure and subject familiarization were expedited by pressurizing first in a sitting position. While the subject was seated in the chair, the head movements (ventral, ventral to dorsal, and lateral) were determined and then the suit was depressurized to allow the subject to stand. A readjustment of helmet tie-down straps was necessary each time a different position was assumed by the subject. (It was impossible to adjust helmet straps while the suit was pressurized to more than 1 psig.) The subject was then readied for the head-rotation test. Straps were adjusted, helmet and suit pressurized to 3.0 psig, and readings were obtained on maximum head deflection. After the head-rotation test, the suit was depressurized, the subject was escorted to the box employed for the standing-flexion test, the suit was pressurized, and this test conducted. The final test was the upper leg, extension backward. After the standing-flexion test, the suit was depressurized to allow the subject to climb off the box and move into position for performing the leg-extension test. The suit was then repressurized and the test measurements made.

RESULTS AND DISCUSSION

Suit weights by components are shown in Table 11-1.

TABLE 11-1 *Suit Weight by Component*

	Suit I		Suit II		Suit III		Suit IV	
	Lb	Oz	Lb	Oz	Lb	Oz	Lb	Oz
Suit	19	6	11	0	9	6	12	0
Helmet	6	4	5	0	6	4	5	0
Gloves	0	12	0	8	1	0	1	0
Boots	5	10	3	6	3	6	3	6
Underwear	1	0	1	0	1	0	1	0
Total weight	33	0	20	14	21	0	22	6

Body Flexion

Some of the tests, particularly test 1, standing flexion, were dependent, to a measurable extent, on the adjustment of the several restraining straps that were an integral part of each pressure suit tested. For the purpose of these tests, constant settings were established as accurately as possible by assessing the fittings of the individual suits on each subject and setting the straps for each posture in a snug setting.

Test 1: *Standing Flexion.* Figure 11-3 shows the suit comparison on the basis of standing-flexion measurement. The greater restriction to body bending by suit I correlates with its greater bulk and weight. Suit I weighed nearly half again as much as the next lighter suit (33 lb) as tested without the back pack. The other suits were similar to each other, suit II weighing 20 lb 14 oz, suit III weighing 21 lb, and suit IV weighing 22 lb 6 oz.

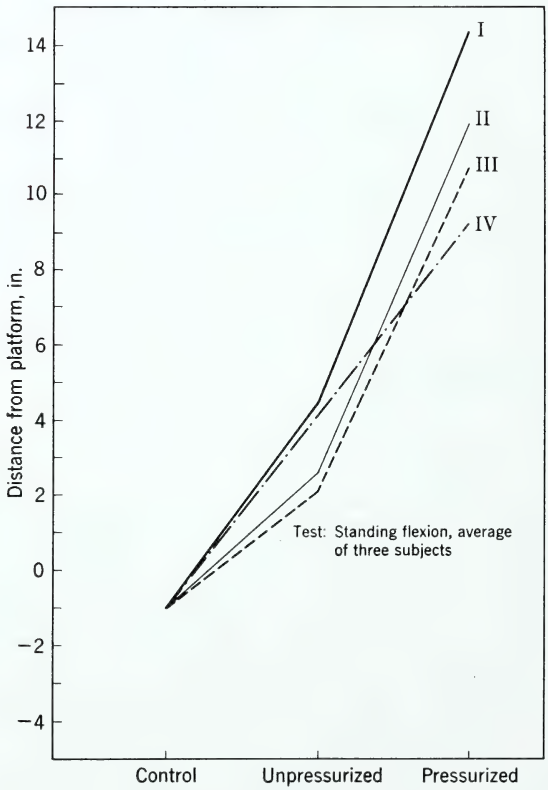


FIG. 11-3 *Standing flexion—suit comparison.*

The overall fit of the suits varied also. Suit IV, for instance, was snug in the chest and abdominal areas when compared with the other suits. This was true for all subjects of similar physical characteristics. With suit III considerable ballooning occurred around the buttocks. This excess material acted as a reservoir for entrapped air. When the subject bent down, the excess material in the area of the buttocks literally "popped out." It was then difficult for the subject to stand as straight as he had been standing originally. The excess material had to be forced back into

folds before the subject could stand straight again. This phenomenon was particularly noticeable in suit III, although all suits displayed similar characteristics in this area. Figure 11-4 shows this ballooning effect in an experimental full-pressure suit.

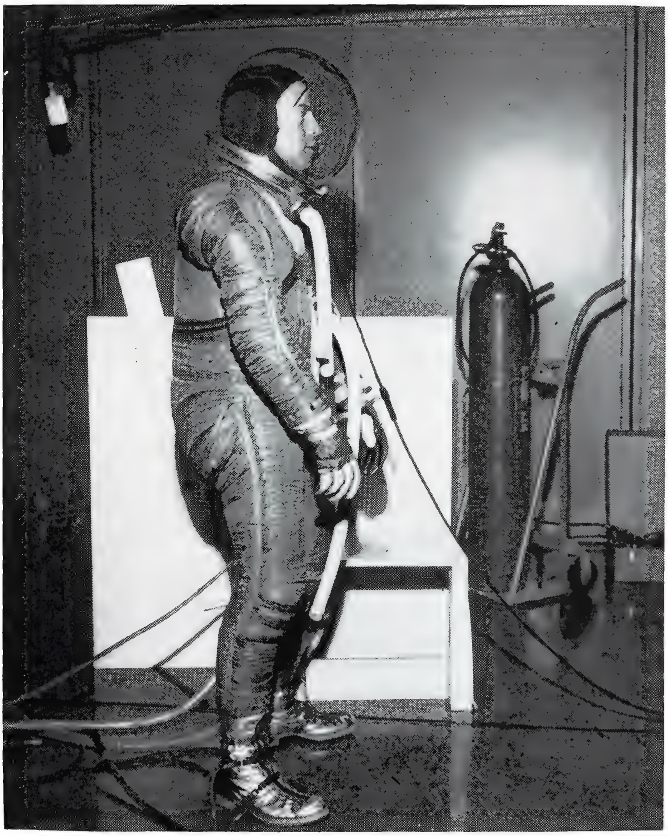


FIG. 11-4 *Experimental full-pressure suit, pressurized to 3.5 psig, illustrating ballooning of fabric in the area of the buttocks.*

The type of "tie-down" or "hold-down" strapping for the helmet contributed to the attenuation of motion. The tie-down strap, or straps, that ran under the crotch effectively restrained the suit fabric until movement was attempted. The subsequent extension and compression of the suit tended to unfold the excess material and refold it in a different fashion. On suits III and IV the single strap running under the crotch could be slipped during the bending operation, whereas the tie-down straps of suits I and II were attached to frontal areas of the suits. Thus, when the suits were pressurized, movements were restricted more in

suits I and II than in III and IV. With suits in the pressurized state, as shown in Figure 11-3, the tighter fit of suit IV allowed greater deflection than did suit III.

Test 2: Head Flexion, Ventral. Figure 11-5 shows graphically the

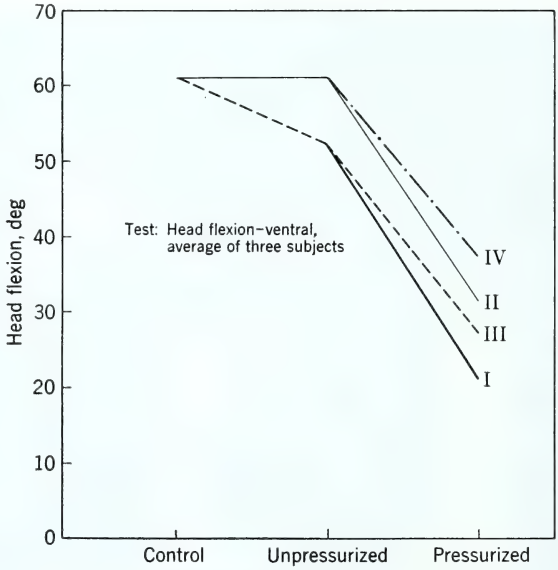


FIG. 11-5 *Head flexion, ventral—suit comparison.*

freedom of ventral head motion in the various suits. The degree of freedom corresponded with the size of neck ring provided on the suit. Suits I and III required a helmet with a larger ring than suits II and IV. The larger neck ring impaired head movement, not only during the ventral motion, but also during the ventral to dorsal and the lateral motions. In the unpressurized condition the small neck ring provided no detectable attenuation of ventral motion. With the suits utilizing the larger neck seal ring an attenuation of 9° was noted, the deflection being cut from 61 to 52° on both suits I and III.

Pressurization of the suit further restricted the head movement. Again the bulk and stiffness of suit I provided more hindrance than the other suits. The larger neck ring also contributed to this condition, since suit III attenuated movement more than suit II. Although suits II and IV utilized the same helmet, suit IV provided less restriction, apparently because of the construction of the suit immediately below the neck ring. Suit II provided an excess of material in the frontal neck portion of the suit in order to facilitate entry into and exit from the suit. These provisions were made in the rear section of the neck on suit IV. The ballooning of this neck area on suit II hindered ventral motion.

The helmet tie-down cables also were not conducive to any extreme ventral motion. The cables on suit IV allowed the helmet to slide along them more readily than did the similar cables of suit II. The net result was that suit IV provided the least attenuation.

Test 3: *Head Flexion, Ventral to Dorsal.* The total angle through which the head and neck may be flexed and extended in the saggital plane is shown in Figure 11-6. Here the size of the neck ring was the dominant factor in determining the restriction provided by the unpressurized suits. With the suits pressurized, however, the dominant factor was the tie-down cable attachment. Both suits III and IV employed similar connections, as was true of the connections of suits I and II. The tie-down accommodations and the previously mentioned entry and exit provisions combined to provide during pressurization more attenuation in suit II than in suit III. Because of a better "fit" than other suits, suit IV was the least restrictive. Suit I, again on account of its bulk, was the most restrictive. As shown in Figure 11-6, the pressurized pressure suit restricted ventral to dorsal

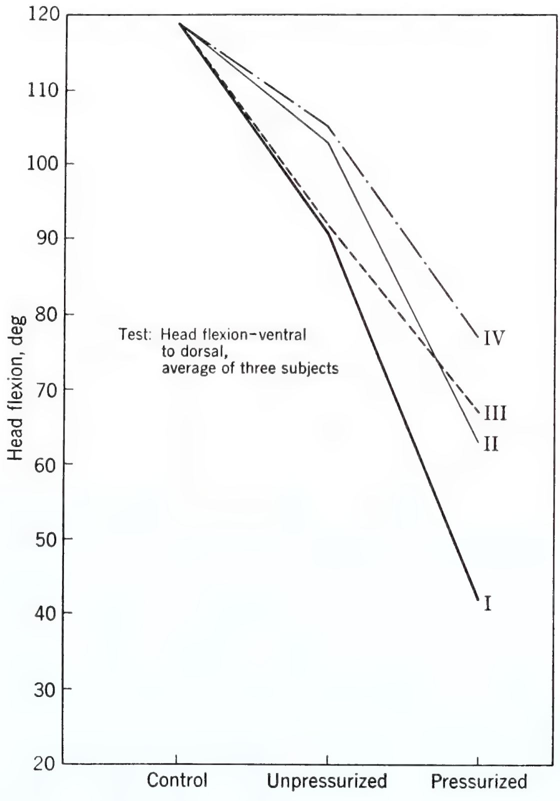


FIG. 11-6 *Head flexion, ventral-dorsal—suit comparison.*

movement to approximately one-half that possible with loose-fitting coveralls.

Test 4: *Head Flexion, Lateral.* Lateral head movement (bending the head from side to side in the plane of the shoulders) was practically eliminated in a pressurized suit, as shown in Figure 11-7. The neck ring,

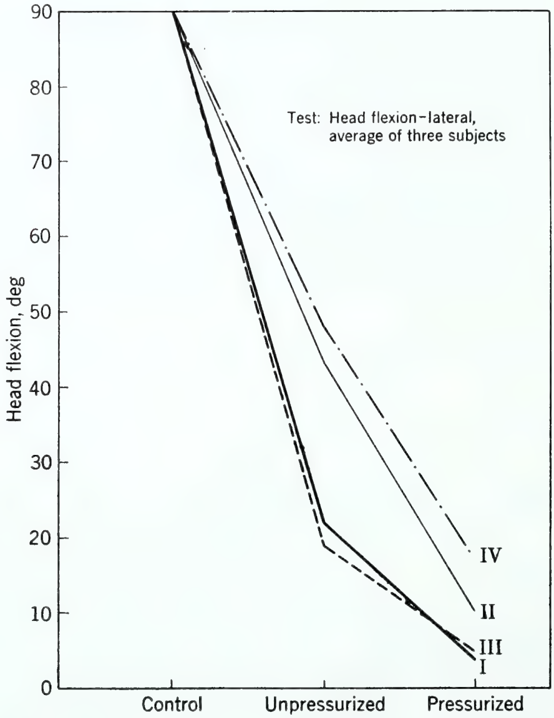


FIG. 11-7 *Head flexion, lateral—suit comparison.*

of basically identical design in all four suits tested, extended outside the ventral neck lines of all four subjects and nearly touched the shoulders. With the head held erect, the ring was immediately above the trapezius muscle. When the suit was unpressurized, the fabric provided flexibility, since the neck ring could move only slightly. The outside diameter of the neck ring on suits I and III was approximately 11 in., whereas the ring employed on suits II and IV had an outside diameter of 9¾ in. The smaller ring, on suits II and IV, allowed more lateral head movement than the larger ring.

Test 5: *Head Rotation.* Figure 11-8 illustrates the results obtained on the four pressure suits for this test. Rotation of the head on the vertebral axis in a horizontal plane is the head movement that permits

maximum use of the window area designed into pilot compartments. Each of the four suits permitted 360° rotation of the helmet within the ball-bearing neck ring. However, limitation of motion resulted from the interaction of several factors, among which were (1) the sizing and fitting of

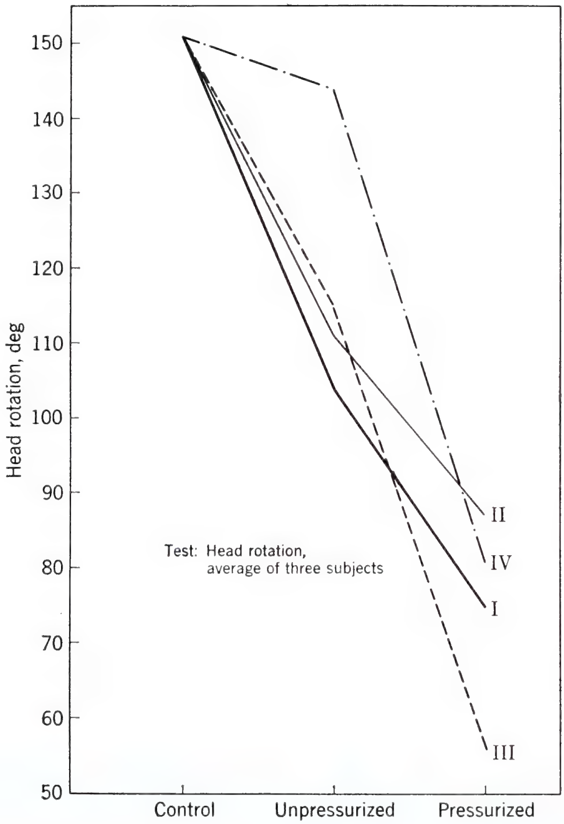


FIG. 11-8 *Head rotation—suit comparison.*

the suit to the individual subject, (2) the location of the ring as determined by the distance from the subject's shoulder to the top of the head, (3) the flexibility of the suit fabric, and (4) the condition of the ball-bearing race in the neck seal. The helmet face-seal adjustment itself was also responsible for some variation.

Test 6: *Upper Leg, Extension Backward.* Very little restriction, as shown in Figure 11-9, was imposed on the subjects by the unpressurized suit during the backward extension of the leg. In fact, more favorable weight distribution resulted, since the helmet forced the head away from the wall. The backward leg movement in the unpressurized suit was, in

most cases, not hindered appreciably, and for suit IV the movement was made easier. During the pressurized operation suits I and III tended to arch the back of the subject forward, placing the subject in an unnatural position for backward leg movement. Suits II and IV allowed

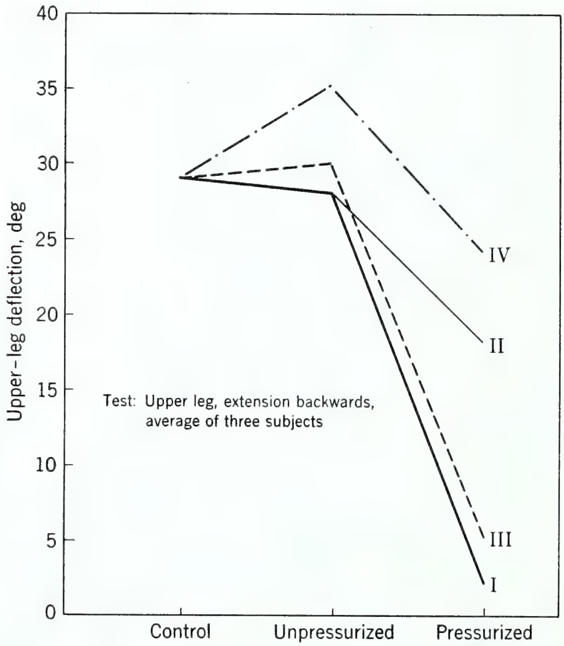


FIG. 11-9 *Upper leg, extension backward—suit comparison.*

the subject to stand in a nearly erect condition, which allowed more favorable balance. Fortunately, the backward extension of the leg is not a characteristic pilot motion at the present time, but it does provide a measure of the flexibility of the suit.

Although no numerical interpretations were made, it was observed that in the pressurized condition it was very difficult to walk in the suits. The subject's balance was unstable as the feet were shifted. Walking was accomplished by keeping the legs stiff and shuffling the feet to and fro by weaving the body from side to side and removing the weight from the foot to be moved.

This study of six types of body flexion was intended to provide operator-mobility measurements for use (1) by engineers concerned with vehicle design, (2) in human factors considerations of man-vehicle relations, (3) in assisting pressure-suit designers, and (4) as a basis for comparison of pressure suits for use in specific vehicles. The study

demonstrated that vehicle designers must take into consideration the degree of restriction of human movement that is introduced by the addition of the pressurized suit. The results indicate that, in fact, a considerable decrement in mobility is caused by the suit even without pressurization.

BIBLIOGRAPHY

1. Dusek, E. R., and W. H. Teichner: The Reliability and Intercorrelations of Eight Tests of Body Flexion, *Quartermaster Research Eng. Command Tech. Rept.* EP-31, May, 1956.

Part D

HUMAN FACTORS IN HIGHWAY SAFETY

Chapter 12

THE ROLE OF HUMAN ENGINEERING IN HIGHWAY SAFETY

*Ross A. McFarland**

THE PURPOSE OF THIS CHAPTER IS TO OUTLINE THE PAST, PRESENT, AND future applications of human factors research in highway safety, with special reference to the design of vehicles and the control of the environment. Thus far the human factors approach has not been used to greatest advantage in this important area.

Human factors, as an organized discipline, may be said to have arisen within the aircraft industry and is to a great extent still centered there. In a recent survey of human engineering activities, it was found that virtually all manufacturers of airframes, missiles, helicopters, and aircraft components either have human factors programs within the company or utilize the services of specialists in this field. The average size of these company groups is 10 persons, with a range of 1 to 34 [10].

These human factors programs have contributed substantially to aviation safety and to survival in aircraft crashes. Less effort has been devoted by the automotive industry to human engineering design. It is, however, encouraging to note that most of the major manufacturers here and abroad are beginning to develop human engineering programs. The human factors research in the aircraft industry could well serve as a model for human engineering programs in other industries. There is little doubt that the application of human factors knowledge to problems of highway safety would appreciably reduce highway accidents and the injuries and deaths resulting from these accidents. The need for such a program is outlined below, and some of the general approaches and techniques available at this time are discussed.

* Harvard School of Public Health, Boston, Mass.

THE PROBLEM OF HIGHWAY SAFETY

Accident Statistics

On the basis of data from those countries for which accurate statistics are available it has been found that accidents are responsible for more deaths than any single illness except cancer and cardiovascular disease. The largest single category of these accidents involves four- and two-wheeled vehicles (see Figure 12-1), and it is estimated that every year over 100,000 persons lose their lives on the world's highways [25].

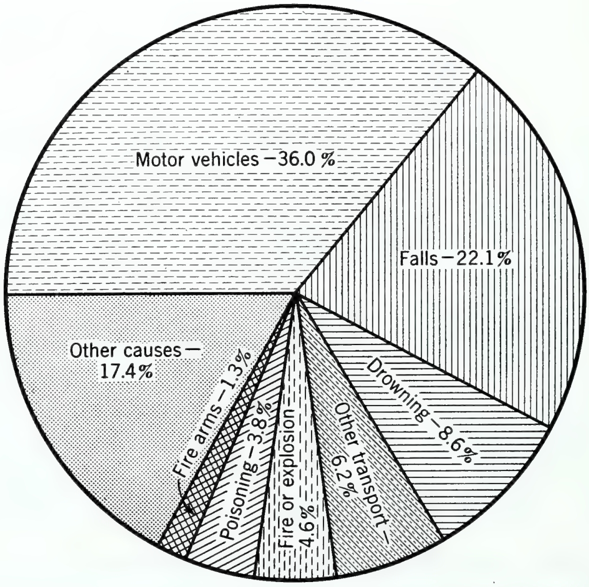


FIG. 12-1 *The distribution of accidental deaths by class of accident in 19 countries representing 553 million people. (Based on data from What Is an Accident? World Health, vol. 14, no. 2, p. 14, 1961.)*

In the United States the accident problem is more serious than in many other parts of the world. Present figures indicate that every year in this country about 93,000 people are killed in accidents (see Figure 12-2), over 400,000 are permanently disabled, and an additional 9.4 million sustain injuries disabling them for 1 day or more after the accident. If the criterion of injury is expanded to cover all those cases in which medical attention is received, then about 47 million persons are so affected each year—over one-quarter of the entire United States population [30]. The cost of all accidents, including wage losses, medical fees and hospital expenses, insurance costs, and property damage, amounts to over 13.6 billion dollars a year [1], roughly 3 per cent of the entire

national income. Accidents are the leading cause of death for all persons between the ages of one and thirty-five, and in the older age groups accidents are outranked only by cancer, heart disease, and the so-called degenerative diseases.

Present figures indicate that in the United States motor-vehicle accidents are responsible for some 38,000 deaths (see Figure 12-2)

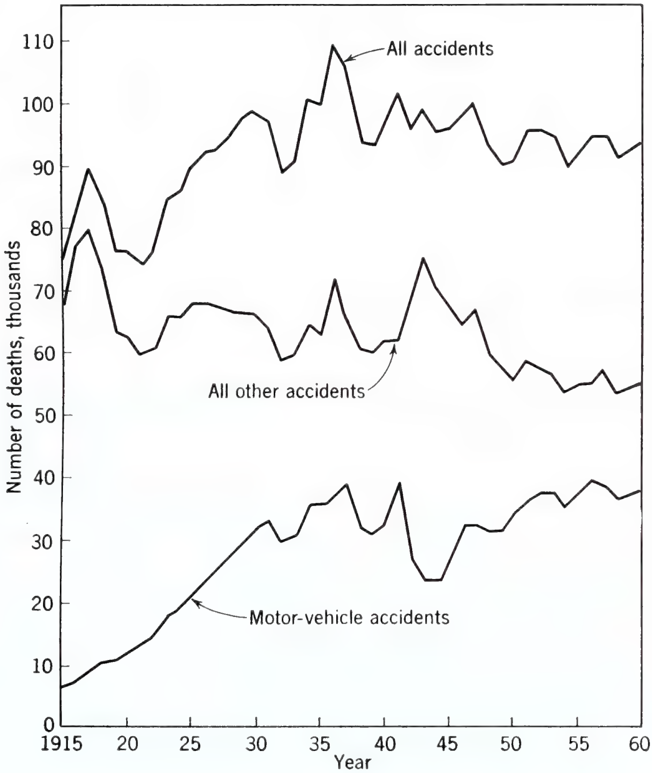


FIG. 12-2 *The number of accidental deaths in the United States. (Based on data from the Accident Prevention Division, U.S. Public Health Service.)*

annually, or about 40 per cent of the total deaths for all accidents, and for about 1,350,000 injuries that are disabling beyond the day of the accident. The annual cost of these motor-vehicle accidents is over 5 billion dollars. It is of interest that, while mileage per death rates have shown an improvement, motor-vehicle fatality rates based on the total population of the United States have remained at essentially the same level for the past decade. Apparently one may expect to drive farther in a given year without being killed, but one is just as likely as before to be killed within that year [12].

At present there are in the United States some 87 million licensed drivers and 74 million vehicles. In 1960 these vehicles were driven a total of 730 billion miles on 3.5 million miles of road. There is now one vehicle for every 2½ persons in the country, 21 vehicles for every mile of road, and 1 mile of road for each square mile of land area [6]. Since the general trend is toward a steady increase in the number of drivers and vehicles and in the length of time spent in vehicles, there seems to be little prospect of a decrease in motor-vehicle accidents and the ensuing deaths, injuries, and financial losses unless special steps are taken. At current rates, 1 million persons will be killed on United States roads during the next 20 years [4]. Estimates indicate that in 1961 motor vehicles will travel about 750 billion miles and that by 1981 this figure will be increased to about 1,300 billion vehicle-miles [6], with the possibility of a corresponding increase in accidents, injuries, and deaths.

The measurement of the number of deaths caused by motor-vehicle accidents is less meaningful than the measurement of how much "lifetime" has been lost prematurely. The years of "expected life lost" can be estimated by multiplying the average life expectancy at each age level, taken from the vital-statistics reports of the United States, by the number of motor-vehicle accident deaths. In 1957 this figure amounted to approximately 1.5 million years. The largest percentage of "life years lost" occurred in the age group between fifteen and twenty-five years, and approximately one-half the total occurred in the group between fifteen and thirty years of age (see Figure 12-3).

The number of male deaths on the highway exceed female deaths by almost 3:1, and the disproportion is most marked in the age range of twenty to forty years, in which male deaths outnumber female deaths by about 6:1. It is also noteworthy that about one-quarter of the pedestrian fatalities are males over sixty-five (see Figure 12-4). Nonfatal injury from motor-vehicle accidents is also higher for young males. Of the 4.7 million estimated to be injured annually, more than one-fourth are in the fifteen- to twenty-four-year age range. More than one-half of all persons requiring hospitalization for injuries are those injured in motor-vehicle accidents [12].

Accidental trauma is a major problem in the armed services as well as in civilian life. For example, during World War II the United States Army, for the first time in its history, reported more deaths due to accidents than to disease. Every fifth death was related to nonbattle trauma, whereas every eighteenth death was related to disease. In the Korean conflict more than half the hospitalized casualties had been injured in accidents rather than by enemy action. Of these casualties, 70 per cent had been involved in motor-vehicle accidents. The frequency

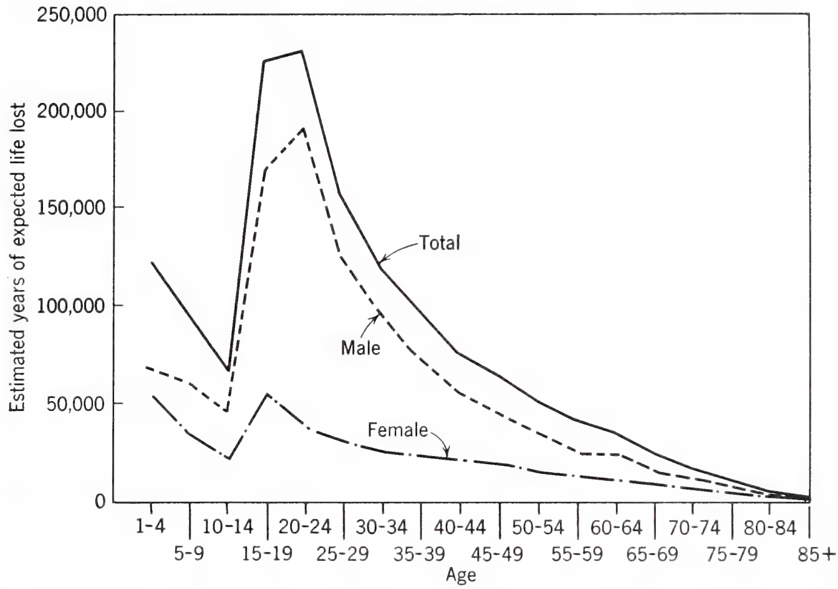


FIG. 12-3 *The number of life years lost as a result of motor-vehicle accidents in 1957 (1,441,000 accidents). (Based on data from "Vital Statistics of the U.S.," U.S. Public Health Service, 1957.)*

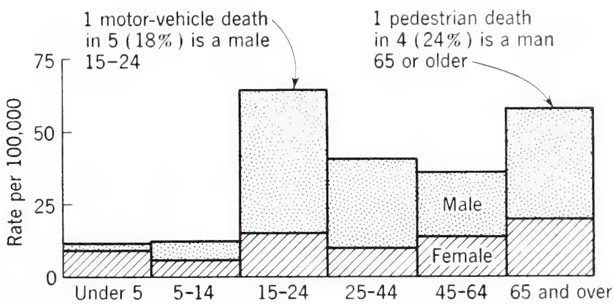


FIG. 12-4 *Motor-vehicle fatalities by age and sex, 1957. (Based on data from the Accident Prevention Division, U.S. Public Health Service.)*

of motor-vehicle fatalities in all three branches of the military approximates 2,000 cases annually, and accidents now rank as the leading cause of man-days lost.

The Epidemiologic Approach to Highway Accidents

It has been demonstrated in recent years that the study of the causes of accidental trauma can be materially aided by the use of techniques

originally evolved for the investigation of communicable diseases. It has not been generally appreciated that accidents exhibit some of the same biological and physiological interrelations that disease processes manifest. When the prevalence and incidence of accidental injuries have been analyzed in a standard epidemiologic manner, it has been shown that accident distributions, like disease distributions, show characteristic variations. Accidental injuries, too, occur at different rates and according to different agents and circumstances among different groups of the population.

Accidental injury, like any disease, arises in the presence of a susceptible victim, the driver in this case, an inciting agent, here the vehicle, and a predisposing environment. The epidemiologic approach to accidents involves a study of their distribution in terms of who had them and when, where, and how they occurred. This descriptive phase of epidemiology is based on sources such as vital statistics, official reports, and various public records. In assessing the magnitude and scope of the accident problem, the relative significance of the various types of accidents and the circumstances contributing to the accident must be considered. This aids in the identification of segments of the population in which accidental injury is disproportionately high or in which specific types of accidents are important. This step also serves as a starting point for the development of causal hypotheses to explain associations apparent in the data and for analytic, detailed studies to test these hypotheses. When the basic physical, physiological, and psychological characteristics of the driver are associated with the vehicle and a given set of environmental conditions, information that may aid in the understanding and prevention of accidents can be obtained. Factual information of this kind can be discovered only through carefully controlled experimental studies, epidemiologic surveys, and statistical analyses.

The general epidemiologic principles of controlling motor-vehicle accidents involve, singly or in combination, (1) reducing the susceptibility of the driver, (2) making the vehicle less hazardous, and (3) modifying the environment in order to lessen the possibility of adverse driver-vehicle interactions. Although much of the essential information about highway accidents remains to be derived, some factual data have been obtained about each of these areas [12].

The characteristics of the driver have so far shown only a limited association with accident frequency. Evidence indicates that young drivers as a group have higher accident rates than would be expected if age were of no significance. Also, changes in the driving skill of older drivers show correlation with age. Moreover, young males seem to have many more accidents than do young females, although the relative driving-exposure time for the two groups is not known. It is well estab-

lished that boys drive more than girls, but it is not known how much more. The major problem is that of determining to what extent lack of driving experience at any age is responsible for accidents and to what extent accidents may be attributed to the immaturity and the attitudes characteristic of youth.

Driver training is another factor often said to be related to accident frequency. However, although trained drivers do have better records than the untrained, there is some doubt as to whether this is due more to the training itself than to the differences in personality and adjustment behavior between those who elect training and those who do not. Similarly, the concept of accident proneness is of only limited utility. Some people will always appear "accident-prone" on the basis of chance alone, and others will have more accidents simply because they are much more frequently exposed to risk.

More fruitful is the investigation of the relation of personal and social adjustments to various aspects of life, including driving. There is some indication that a man does "drive as he lives"; it has been demonstrated that those known to such agencies as courts, collection and credit organizations, and social-welfare groups are much more likely to have repeated accidents than those not known to such agencies. Driver attitudes are a characteristic of the utmost importance, but there is at present little information about the attitudes characteristic of safe drivers as contrasted with unsafe drivers.

Various temporary conditions of the driver have also been shown to be related to accident frequency: (1) the deterioration of skill with fatigue; (2) the disorganization of skill with various emotional states; (3) the effects of alcohol; (4) the effects of various diseases; (5) the influence of drugs and medications [12].

Many characteristics of the driver are relevant to highway safety, but it must be remembered that the vehicle and certain aspects of the environment are generally more amenable to control than driver characteristics and in some cases show more promise of immediate and specific gains in accident reduction. It is in the control of the vehicle and the driving environment that the human engineer can make his unique and most important contribution to highway safety.

APPLICATIONS OF HUMAN ENGINEERING TO HIGHWAY SAFETY

The Scope of Human Engineering in Highway Safety

One of the primary objectives of human engineering is that of improving safety by designing equipment in terms of human capabilities

and limitations. From the standpoint of human engineering, it is essential that the mechanical design of automotive equipment be compatible with the biological and psychological characteristics of the driver. The effectiveness of the automotive man-machine combination can be greatly enhanced by treating the operator and the automobile as a unified system. Thus, the instruments should be considered as extensions of the driver's nervous and perceptive systems, the controls as extensions of the hands, and the feet as simple tools. In general, any control difficult to reach or operate, any instrument dial of poor legibility, any seat inducing poor posture or discomfort, or any unnecessary obstruction to vision may contribute directly to an accident [14].

When the vehicle driver is viewed as one component of a man-machine system, the general human characteristics pertinent to design are: (1) physical dimensions; (2) capability for data sensing; (3) capability for data processing; (4) capability for motor activity; (5) capability for learning; (6) physical and psychological needs; (7) sensitivities to the physical environment; (8) sensitivities to the social environment; (9) capability for coordinated action; (10) differences among individuals [27]. Quantitative information about these human characteristics must be coordinated with the data on vehicle characteristics if maximum man-machine integration is to be achieved.

Unfortunately, much of the biological and psychological information needed for this purpose is not yet available, although pertinent human factors research is currently being carried out by many organizations. One of the most significant developments consists of the technological advances that have been made in electronics, especially in regard to high-speed monitoring, recording, and computing procedures, which make possible the effective and quantitative study of the human component in relation to the other components of the system and to the output of the total system. The application of digital and analog computers in simulating the performance of aircraft and submarines is one example of the progress made in this area. These techniques may soon be available for the study of the driver-vehicle-roadway system.

In any such approach the driver would be considered a unique servomechanism in the man-machine system that alters the performance output of the system in accordance with the continuous sensory and perceptual input derived from the vehicle's responses, the environment, and the driver's judgments. The functioning of the nonhuman components of the system can be expressed in precise mathematical terms, but it would be difficult to include the human operator in these formulations. Except for a small number of activities, the human operator behaves in a nonlinear fashion; i.e., his output does not necessarily equal the sum of his inputs. The important point for this discussion, however, is that

with the new electronic techniques the performance of the vehicle and the environmental factors relevant to driving could be simulated and operator performance could be closely monitored and objectively measured. This procedure would provide the basis for a precise quantitative study of the human problems in highway transportation.

The initial phase of the application of human engineering to highway safety should be an analysis, in advance of manufacture, of all design aspects involving human behavior, or an operational job analysis. This concept of design is relatively new and unexplored in the automotive field, partly because accident reports generally have failed to identify difficulties in man-machine integration as accident causes. Design failures may be so subtle that those responsible for reporting accidents may not be aware of them, particularly if the personnel are not trained in human engineering. However, if defects are present, it is only a matter of time before some driver "fails" and has an accident [14]. The following is an outline of the chief considerations in the advance analysis of equipment for design faults [13]:

1. Operational job analysis
   a. Requirements of driving
   b. Driver's area in the vehicle
   c. Characteristics of instruments, displays, and controls
2. Blueprint phase
   a. Prediction of performance
   b. Human limitations
   c. Anticipation of errors
3. Mockup stage
   a. Performance of duties
   b. Physical size of drivers
   c. Skill requirements
   d. Age considerations
   e. Interfering structures
   f. Physiological stress
   g. Errors or near accidents

There are many examples of ways in which automotive engineers have already improved safety by designing equipment to compensate for human limitations. These include improved braking and steering systems, lighting, tires, and automatic shifting. However, carefully controlled mockup studies for testing the nature and range of human abilities might aid in making the most of these advances. How fast should the brakes operate in relation to the age, sex, size, and strength of the driver? How much feedback should be provided by the power-steering system? How should the automatic shifting device be designed for greatest standardiza-

tion and ease of operation? If the biological scientist were present during the mockup stages, he might well be able to advise the engineer about the range of abilities of the driving population [13].

The specific features of vehicle design that deserve the attention of the human engineer are: (1) the layout of the workspace, including spatial accommodation and seating; (2) the design and arrangement of controls; (3) the design and arrangement of displays; (4) the design of window areas for maximum vision from the vehicle; (5) the design of the vehicle to provide maximum occupant protection during an accident; (6) the control of those aspects of the physical environment which influence the driver; (7) the prediction of error and the provision of margins of safety.

A series of human engineering evaluations of this sort has been carried out on commercial vehicles [24] and, more recently, on 20 domestic and 22 foreign passenger cars [18]. Some of the findings of these studies are presented in the following discussions of specific applications of human engineering to vehicle design. Further details are presented in Chapter 14, The Operator and Vehicle Design.

The Layout of the Workspace

In order to ensure the proper layout of the workspace, the dimensions of both the operator and the vehicle (Figure 12-5*a* and *b*) must be

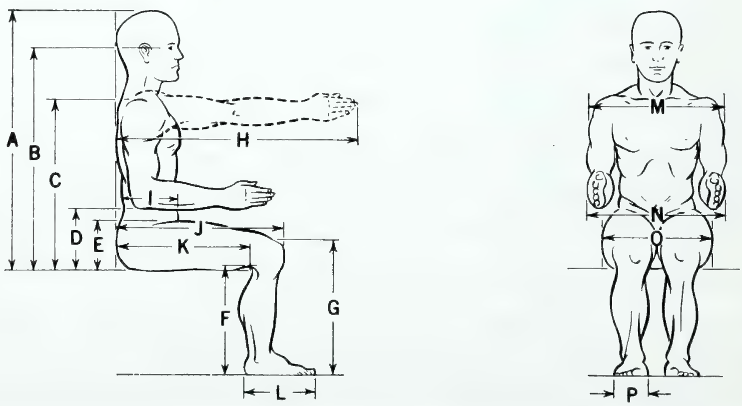


FIG. 12-5*a* *Human body measurements relevant to vehicle design (see also Table 12-2). (After R. A. McFarland and H. W. Stoudt, Human Body Size and Passenger Vehicle Design, SAE Spec. Publ. 142A, 1961. By permission of the publishers.)*

known. The body dimensions necessary for establishing the dimensions of the workspace and for locating the controls are: (1) the maximum

arm reaches attainable without alteration of body position; (2) the extension of these reaches which can be attained by movement of the trunk or body; (3) the eye level of the man in the operating position; (4) the body dimensions in the operating position, i.e., sitting heights and fore-and-aft and lateral measurements at various levels; (5) the leg reaches attainable without alteration of posture [9].

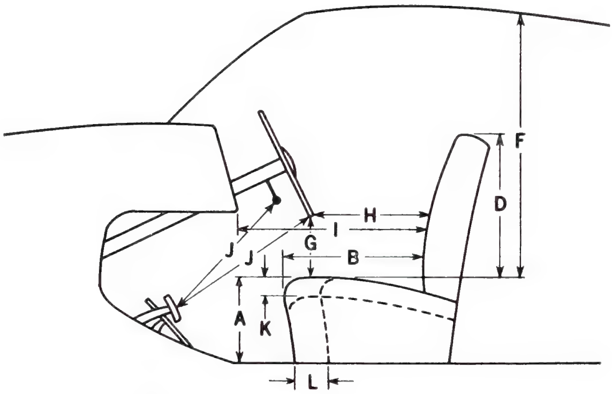


FIG. 12-5b *Some vehicle dimensions related to human body measurements. (After R. A. McFarland and H. W. Stoudt, Human Body Size and Passenger Vehicle Design, SAE Spec. Publ. 142A, 1961. By permission of the publishers.)*

Whereas a great deal of human-sizing data has been obtained for pilots in connection with cockpit design, little specific information is available for motor-vehicle drivers. In surveys conducted by the Harvard School of Public Health, measurements of 32 body dimensions were made on a representative sample of approximately 360 bus and truck drivers in all parts of the country, including 100 champion truck drivers competing in a National Roadeo of driving skill. These data were supplemented by measurements taken on a large sample of military drivers. However, suitable comparable data on the general driving population for the design of passenger vehicles are still unavailable, although approximations of the body measurements of this group have been interpolated from selected anthropometric studies of various segments of the United States population. Admittedly, direct measurements of the driving population are preferable; however, the tentative values presented in Table 12-1 should indicate the true measurements closely enough so that they may be used as interim values with some confidence. Many of the needed body measurements are currently being taken by the National Health Survey of the U.S. Public Health Service on a statistically representative sample of the entire United States population,

TABLE 12-1 *Approximations of Body Measurements Related to Vehicle Design for the General Driving Population*

Body measurements, in.	Male drivers			Female drivers		
	Percentiles			Percentiles		
	5th	50th	95th	5th	50th	95th
Stature	64.1	68.4	72.6	59.5	63.4	67.3
Weight, lb	131	166	216	105	136	190
A.*Sitting height	33.8	36.0	38.2	31.6	33.7	35.6
B. Eye height	29.3	31.6	33.7	27.2	29.3	31.1
C. Shoulder height	21.4	23.3	25.2	19.3	21.1	22.9
D. Elbow height	7.7	9.3	10.9	8.2	9.4	10.9
E. Thigh height	4.8	5.7	6.8	4.9	5.8	6.7
F. Popliteal height	15.5	16.9	18.1	13.9	15.2	16.2
G. Knee height	19.8	21.6	23.5	17.9	19.5	20.8
H. Anterior arm reach	32.0	34.9	37.5	28.5	30.9	33.4
I. Abdomen depth	8.4	10.1	12.4	7.9	9.0	11.1
J. Buttock-knee length	21.6	23.5	25.5	20.7	22.4	24.0
K. Buttock-popliteal length	17.4	18.9	20.8	16.8	18.2	20.0
L. Foot length	9.6	10.4	11.3	8.8	9.6	10.2
M. Shoulder breadth	16.4	17.7	19.4	14.3	15.6	17.5
N. Elbow breadth	15.0	17.4	20.6	13.4	15.0	16.8
O. Seat breadth	13.0	14.4	16.2	13.6	15.1	17.2
P. Foot breadth	3.6	3.9	4.3	3.3	3.6	3.9

* Letters refer to the dimensions shown in Figure 12-5*a*.

and these data should be available within the next year or two. In the meantime, on the basis of existing anthropometric data it has been possible to specify dimensions for certain distances within the driver's workspace to accommodate the middle 90 per cent of the drivers with respect to body size [17, 21].

Table 12-2 gives examples of percentile distributions for three body measurements known to be useful in the design of the driver's workspace. It should be noted that these data are not presented in terms of averages: the use of average values may account for many defects in vehicle design, since arrangements for the person of "average size" might be suitable for only 50 per cent of the operators in a normally distributed group. Provision for 90 or 95 per cent, or any other predetermined proportion, of potential operators requires identification of the correct cutoff points. When arm reach for the operation of controls is under consideration, the cutoff point should be well below the average reach.

In several of the commercial vehicles evaluated, for example, only

TABLE 12-2 *Percentile Distributions of Three Selected Body Measurements of Commercial Truck and Bus Drivers*

Percentile	Sitting height	Knee height	Anterior arm reach
5	34.3	20.1	33.0
10	34.8	20.4	33.5
15	35.1	20.6	33.9
20	35.3	20.8	34.3
25	35.5	21.0	34.7
30	35.7	21.2	34.9
35	35.9	21.3	35.1
40	36.0	21.4	35.3
45	36.2	21.6	35.5
50	36.3	21.7	35.8
55	36.5	21.8	35.9
60	36.7	21.9	36.1
65	36.8	22.1	36.4
70	36.9	22.2	36.6
75	37.1	22.2	36.9
80	37.3	22.4	37.1
85	37.6	22.6	37.4
90	37.8	22.8	37.9
95	38.2	23.5	38.4
	N = 310	N = 301	N = 312
	Range = 33.1–39.0	Range = 19.3–26.0	Range = 30.7–41.7

5 per cent of the drivers could comfortably reach and operate the hand brake. In other vehicles, only 60 per cent of the drivers could be accommodated for knee height between the pedals and the steering wheel. Many tall drivers were unable to adjust their sitting positions to obtain maximum visibility in relation to their instruments and the road ahead [24].

Failure to provide for adequate seat adjustment to allow for variations in human size was frequently reported by the drivers and was evident from the objective studies of work areas. Probably the most striking defect was that the front of the seat could not be lowered to enable shorter individuals to operate the pedals without excessive pressure under the knees. Some of the medical problems frequently observed in truck drivers are believed to be related to poor seat design and to failure to provide adequate shock absorbers [20].

Integrating anthropometric data with the static dimensions of the cab is not always easy, however, since most body measurements are

made under static conditions and driving is a dynamic situation. As a result, static data cannot always be applied without additional study using mockups and subjects representing known points in the distribution of body measurements. For example, when one truck model was evaluated, it was discovered that the taller drivers could not operate the foot brake when the gear lever was in either of the two left positions. It was impossible for a large driver to put his foot on the brake pedal without first shifting gears. The distance between the brake pedal and the bottom of the steering wheel was too short to permit the leg to move high enough to put the foot on the pedal, and the gearshift in either of the left positions was too close to the wheel to allow the leg to slip between. As a result, the foot was trapped on the floor until the driver shifted gears [24]. This defect should have been eliminated in the preproduction stages of the vehicle.

The principle that static dimensions require supplementary techniques involving the dynamic setting of the driver at his task was applied in an additional experiment in which the objective was that of determining the ranges of adjustability in seating and controls for optimal accommodations for individual operators. Subjects representing the 5th, 50th, and 95th percentiles of drivers "drove" an adjustable mockup in a simulated driving task. The interrelations between the various items in the cab were determined by statistical analysis, and recommendations were made for their location and ranges of adjustability. A significant finding of the experiment was that fewer errors were made in the driving task when the settings in the apparatus were those at which the driver was most comfortable [8].

The Design of Controls

If controls were so designed and located that they could be operated easily, accurately, and rapidly, many driver errors leading to highway accidents could be prevented. For example, one serious accident resulted when a driver, while proceeding at high speed in a modern car, shut off his headlights in the belief that he was operating the cigarette lighter. The knobs for these two controls were identical in shape and size and were located near each other on the dashboard. In two truck models the dimmer switch was located directly beneath the foot pedals and close to the steering post. It is likely, therefore, that the driver might operate other switches or equipment while attempting to dim the head lamps. Even when he operates correctly, complex motions and long movement times are required for avoiding the pedals. Adequate shape coding for the accurate identification of knobs and switches by touch and increased standardization in their location would probably aid in

the reduction of accidents caused by the inadvertent operation of wrong controls.

The brake and accelerator pedals also require study and improvement. In some vehicles these pedals are so placed that the driver must make lengthy movements in three directions before the brake can be activated, i.e., up, over, and down. Sometimes the two pedals are of identical design and material, and therefore it is impossible for the driver to distinguish the pedals by touch alone. Undoubtedly this defect has been responsible for many critical situations, especially with drivers new to the vehicle, and has probably caused a number of accidents.

During the operation of pedals, various muscle groups may become acutely fatigued from holding a steady position. The lower right leg and foot, for example, may react slowly in emergency situations after maintaining continuous pressure against the accelerator pedal. In some situations, as on runs in mountainous country, continuous turning and shifting are required. Mechanized or hydraulic aids to steering will continue to reduce the possibility of errors during such maneuvers.

The more important considerations in the design of controls may be summarized as follows: (1) location of controls for ease and accuracy of reaching; (2) direction of movement for greatest efficiency; (3) amount of force that must be applied; (4) rate of movement from point to point; (5) speed and amount of rotary or wrist movement required in wheels or knobs; (6) size and shape; (7) frequency of use; (8) the degree to which the control performs a critical function. A considerable amount of data on these factors is currently available and may be applied to problems of automotive design [15, 16].

The Design of Instruments

A driver can successfully control his vehicle only to the extent that he receives unambiguous information on all pertinent aspects of his task. Sometimes accidents, or operational errors, occur because a driver misinterpreted, or was unable to obtain, information from his displays concerning the functioning of the equipment. The three basic types of indicators supplying the driver with information are: (1) the check display, which indicates whether or not a given condition exists, e.g., a red light to show failure, danger, or faulty operation; (2) the quantitative display, which indicates an exact numerical value to be read, e.g., a speedometer or clock; (3) the qualitative display, which indicates whether, and in what direction, an operation has deviated from a desired level, without presenting precise quantitative data, e.g., a temperature gauge [2]. Many quantitative displays could profitably be changed to

the qualitative type of indicator. Whenever possible, instruments should have the least complex type of display consistent with the degree of precision required.

For best possible efficiency, all instruments should possess the following characteristics: (1) can be read quickly; (2) can be read as accurately as is required—and preferably no more accurately; (3) presents no ambiguity or likelihood of gross reading errors; (4) the information is provided in the most immediately useful form and does not require mental translation into other units; (5) changes in value are easily detected; (6) can be easily identified and distinguished from other instruments; (7) the information is as current as possible; (8) if inoperative, either cannot be read or the operator is properly warned [26]. The specific factors that influence the ease, speed, and accuracy with which instruments can be read include location with reference to the operator, the size, shape, and spacing of the critical detail, and the brightness and contrast of the display. Design specifications for most of these factors have been determined experimentally [2].

The Design of Window Areas

The design of windshields, side windows, and rear-vision mirrors is an important area for human engineering research. There has been developed an instrument that precisely delineates the area outside the vehicle that is seen by the driver and thus permits accurate evaluations of the relative efficiency of various configurations of windshields, windows, and supporting structures [29].

One important problem is that of evaluating tinted windshields, which have been widely accepted in the United States. There are certain advantages in reducing the glare from bright sunlight and aiding in the control of temperature within the car. However, since the National Safety Council reports that accident rates are three times higher at night than in the daytime, this design feature deserves careful consideration. In the evaluation of tinted windshields the age of the driver is an extremely important factor, since it has been established that with increasing age there is a decrease in the retina's sensitivity to light under low levels of illumination. If during night driving this decrease in light sensitivity is combined with the reduced transmission of light through tinted glass, a serious safety problem may arise. (Ordinary safety glass absorbs 12 per cent of the light, whereas tinted glass absorbs 28 per cent.)

Studies have indicated that at any age the effect of the tinted glass at night is a reduction in the visibility of lights and other objects near the threshold of perception. Figure 12-6 illustrates these findings. The

first portion of the curves shows the course of the sensitivity threshold for groups of subjects in the age range of sixteen to eighty-nine years during 40 min of dark adaptation. The slight rise in the thresholds at 41 min resulted from the introduction of clear windshield glass; the substantial rise at 46 min followed the viewing of the test lights through

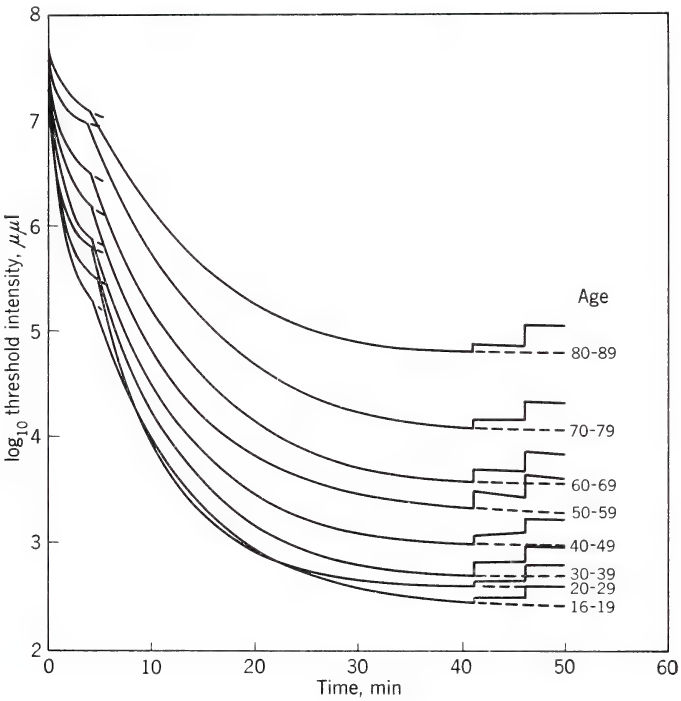


FIG. 12-6 *Dark adaptation as a function of age and tinted windshield glass. Age range = 16–89; N = 240. (After R. A. McFarland and R. G. Domey, Experimental Studies of Night Vision as a Function of Age and Changes in Illumination, Highway Research Board Bull., vol. 191, pp. 17–32, 1958. By permission of the publishers.)*

tinted windshield glass [19]. In these experiments, age was shown to be highly correlated with dark-adaptation thresholds at time intervals of 2 to 40 min. The correlations ranged from .71 at the second minute to .84 at the 40th minute. They are among the highest correlations reported between age and biological function [23].

Crash-injury Protection

Research by De Haven, Stapp, Ryan, and others on the ability of the human body to withstand sudden deceleration and impact forces has

indicated that the force of many fatal accidents may be within the physiological limits of survival [28]. It was found, for example, that a human being can withstand high decelerative forces without injury if he is properly restrained. On the other hand, a speed of 15 mph can cause death if the momentum of the head is not checked during rapid deceleration of the vehicle. If a 10-lb object, the approximate weight of a human skull, fell 1 ft and struck an area 1 in. square, the deformation would be slight, whereas, if under the same conditions an object 1 cm square were struck, a puncture fracture would result.

Controlled research in this area is defining not only the tolerances of the human body to force but also the injury-producing variables within a decelerating vehicle. Automotive engineers can contribute to the reduction of injuries and fatalities by designing the vehicle according to the findings of such studies. Some protective measures have been developed, such as stronger door locks, recessed steering-wheel hubs, and various types of dashboard padding.

Control of the Physical Environment

In some instances the control of highway accidents depends on the regulation of the interaction between the driver and the internal and external environments of the vehicle. The more important physical variables of the environment are temperature, humidity and ventilation, noise, vibration, and carbon monoxide. Extremes of any of these variables will result in discomfort and lowered efficiency of the driver. The ranges of comfort and discomfort and the extremes that may contribute to accidents have been reasonably well defined [22]. Certain aspects of highway design and the influence of weather conditions on driving safety are not considered here.

Temperature, Humidity, and Ventilation. A change in any one of these variables will affect the other two in so far as they influence comfort. Hence, all three variables must be considered as a unit.

The ventilation problem is concerned with the amount of fresh air and air movement provided in the vehicle. Low rates of air movement produce feelings of stuffiness and fail to prevent the accumulation of body odors and tobacco smoke. Optimally 35 to 40 ft^3/min of fresh air should be supplied per individual at velocities of between 20 and 60 ft/min in order to maintain an atmosphere of freshness without creating undesirable drafts. It should be remembered that increased air velocities tend to have a cooling effect, which may be advantageous in summer but which may cause discomfort in winter. Poor ventilation causes lethargy and sleepiness and can contribute to accidents.

Composite curves have been developed to show the various comfort

zones and the typical subjective responses to heat, cold, and humidity [5]. The optimum temperature range is 68 to 72°F (dry bulb) for winter and 74 to 78°F for summer. A relative humidity of 30 to 70 per cent is acceptable in these zones. Excessively high humidities cause discomfort at high temperatures and result in feelings of coolness or dampness at low temperatures. When humidity is below 15 per cent, the membranes of the eyes, nose, and throat become uncomfortably dry [15]. Experiments with industrial workers have shown that human performance deteriorates significantly at effective temperatures of about 83° (the equivalent of 83°F at 100 per cent humidity) and above [11]. Many of the skills that show deterioration are directly related to safe driving.

Noise. In most passenger automobiles noise levels are sufficiently low to require few design changes for driver health and safety. However, the amount of noise produced by some commercial vehicles presents a serious problem. In seven current truck models tested for overall noise levels within the cab it was found that the range was 48 to 89 decibels (db) with the cab closed and the engine idling [20]. In those models that were road-tested under normal operating conditions the second- and third-gear noise levels were between 80 and 95 db, and in fourth gear the noise level was just below 90 db [20]. Callaway's octave-band analysis of the noise outside trucks revealed that heavy trucks reached a peak of 93 db in the 75- to 100-cps frequency band, dropping to 80 db in the 600- to 4,800-cps band [3]. Measurements made on a passenger automobile during the same investigation showed a maximum of only 64 db in the 150- to 300-cps band. The wide difference in noise levels between the quietest and the noisiest trucks indicates, however, that some designers have been able to make improvements in noise control.

Vibration. Vibration in motor vehicles leads to discomfort and fatigue in driving operations, and there is increasing evidence that serious physical disorders may result from protracted exposure to excessive vibration. Even under less extreme conditions, visual acuity is affected and certain reflexes are diminished. Determining the physiologic limits of human tolerance to vibration is of great importance; especially needed are medical and engineering studies of the effects of vibration beyond the tolerance limits for health, safety, and efficiency.

The subjective reaction of the driver varies with the frequency and amplitude of the vibration. In general, the higher the frequency, the lower the amplitude must be to maintain a given level of comfort. Also, as frequencies increase, the effect of vibration tends to become localized, whereas with low frequencies the whole body may be involved. Most of the frequencies encountered in motorized equipment are low, i.e., 8 cps or less. Recent experiments by Hornick et al. [7] on the effects of vibration in ground vehicles should be consulted.

Carbon Monoxide. Carbon monoxide poisoning is an ever-present possibility in the operation of motor vehicles. The problem is becoming increasingly serious because of the amount of smog and the concentration of idling vehicles in metropolitan areas. Small amounts of carbon monoxide are absorbed rapidly by the blood stream, resulting in an oxygen deficiency that may at first be unnoticed by the individual. The initial reaction to carbon monoxide poisoning consists primarily in lowered attention, difficulty in concentration and retention, slight muscular incoordination, sleepiness, and mental and physical lethargy. These effects could easily contribute to an accident. One of the earliest demon-

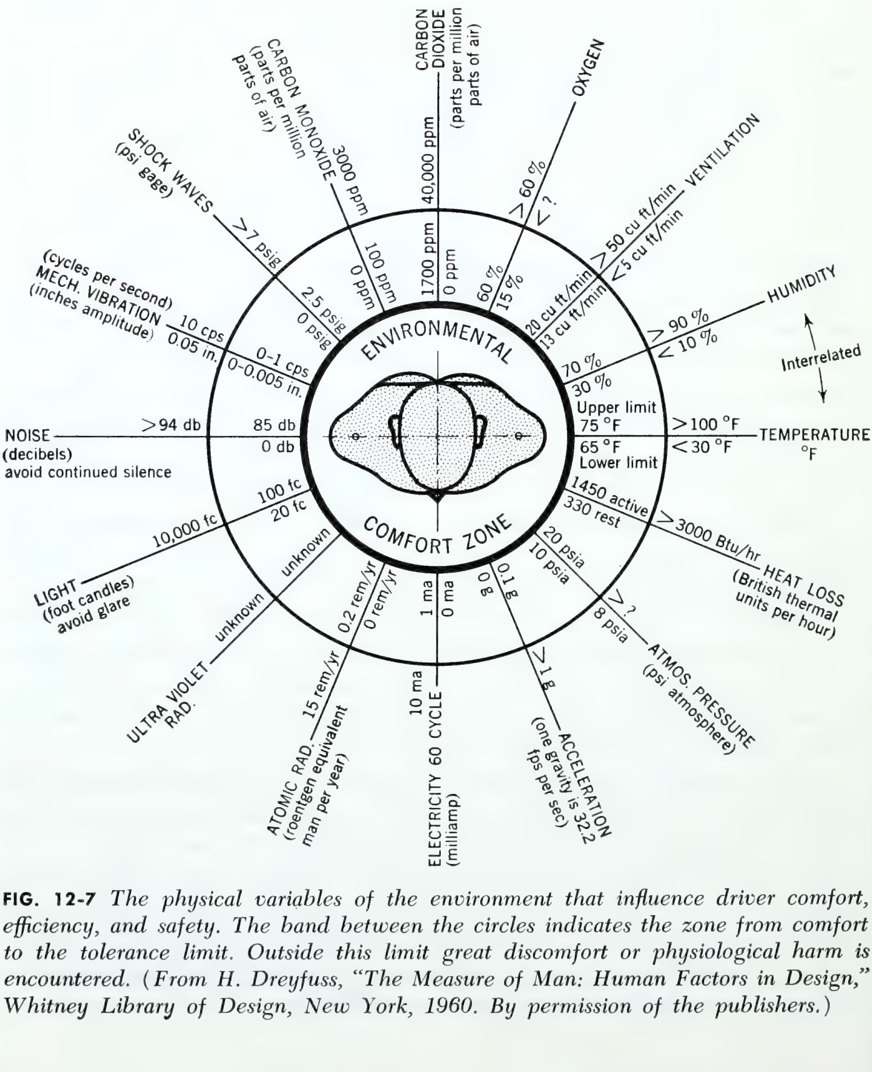


FIG. 12-7 *The physical variables of the environment that influence driver comfort, efficiency, and safety. The band between the circles indicates the zone from comfort to the tolerance limit. Outside this limit great discomfort or physiological harm is encountered. (From H. Dreyfuss, "The Measure of Man: Human Factors in Design," Whitney Library of Design, New York, 1960. By permission of the publishers.)*

strable effects of carbon monoxide is the reduction of the sensitivity of the eye under low illumination. For example, chronic cigarette smokers are known to have 4 to 8 per cent carboxyhemoglobin in their blood, an amount that corresponds in effect to an altitude of about 7,000 ft. If additional amounts are received from the exhaust system of the vehicle, a dangerously high level of oxygen deficiency may result [16]. In motor vehicles dangerous concentrations of carbon monoxide usually result from leaks in the exhaust system and from openings in the floor boards or other parts of the vehicle body. It is desirable to limit the amount of carbon monoxide to 0.003 per cent [16].

A schematic interpretation of the various stimuli encountered in driving is given in Figure 12-7 in terms of the limits for comfort, discomfort, and injury [22]. Because of the interdependence of the variables, these values should not be considered rigid standards. The difficulty in giving a single numerical value as a standard is exemplified by the limits indicated for carbon monoxide and vibration. The generally accepted maximum value for prolonged exposures is 100 parts per million (ppm) of carbon monoxide in circulating air. This amount may be satisfactory for a person sitting quietly at sea level, but it must be smaller if the amount of exercise, the degree of ventilation, or the length of exposure is increased. The permissible amplitudes indicated for vibration apply only to frequencies of 8 cps or less; however, frequencies higher than this are probably not a major factor in automobile operation. Linear scales or straight-line functions are not so adaptable to the study of human variables as are nomograms, in which one can ascertain the effects that altering any one variable has upon the other variables.

SUMMARY

In this chapter the magnitude of the highway-safety problem was spelled out in terms of deaths, injuries, and economics. Various approaches to the problem were discussed, and the ways in which the human factors specialist can contribute to highway safety were defined. It was concluded that his main contribution lies in designing the vehicle and controlling the environment in accordance with the characteristics of the driver. Specific applications of human engineering include the layout of the driver's workspace, design of seats, controls, instruments, and window areas, protection from injury during a crash, and control of those aspects of the physical environment that influence driver behavior. Emphasis was placed on the unity of the driver-vehicle-environment system.

BIBLIOGRAPHY

1. Accident Facts—1961 Edition, National Safety Council, Chicago, 1961.

2. Baker, C. A., and W. F. Grether: Visual Presentation of Information, *USAF WADC Tech. Rept.* 54–160, August, 1954.

3. Callaway, D. B.: Measurement and Evaluation of Exhaust Noise of Over-the-road Trucks, *Trans. SAE,* vol. 62, pp. 151–162, 1954.

4. Conference on Transportation Research, *Natl. Acad. Sci.–Natl. Research Council Publ.* 840, 1960.

5. "Heating, Ventilating, Air Conditioning Guide 1960," American Society of Heating, Refrigerating, and Air-conditioning Engineers, New York, 1960.

6. Holmes, E. H.: Highway Transportation, in U.S. Transportation Resources, Performance, and Problems, pp. 1–86, *Natl. Acad. Sci.–Natl. Research Council Publ.* 841-S, 1961.

7. Hornick, R. J., C. A. Boettcher, and A. K. Simons: The Effect of Low Frequency, High Amplitude, Whole Body, Longitudinal and Transverse Vibration upon Human Performance, Bostrom Research Laboratories, Milwaukee, July, 1961.

8. Kephart, N. C., and J. W. Dunlap: Human Factors in the Design of Vehicle Cab Areas, Occupational Research Center, Purdue University, Lafayette, Ind., 1955.

9. King, B. G.: Measurements of Man for Making Machinery, *Am. J. Phys. Anthropol.,* vol. 6, pp. 341–351, 1948.

10. Kraft, J. A.: A Follow-up Survey of Human Factors Research in Aircraft, Missile, and Supporting Industries, *Human Factors,* vol. 1, no. 1, pp. 23–25, 1958.

11. Mackworth, N. H.: Researches on the Measurement of Human Performance, *Med. Research Council Spec. Rept. Ser.* 268, 1950.

12. McFarland, R. A.: The Epidemiology of Motor Vehicle Accidents, *J. Am. Med. Assoc.,* vol. 180, pp. 289–300, 1962.

13. ———: Human Engineering and Automobile Safety, *Clin. Orthopaed.,* vol. 9, pp. 260–276, 1957.

14. ———: Human Engineering, A New Approach to Driver Efficiency and Transport Safety, *Trans. SAE,* vol. 62, pp. 335–345, 1954.

15. ———: "Human Factors in Air Transport Design," McGraw-Hill Book Company, Inc., New York, 1946.

16. ———: "Human Factors in Air Transportation—Occupational Health and Safety," McGraw-Hill Book Company, Inc., New York, 1953.

17. ———, A. Damon, and H. W. Stoudt: Anthropometry in the Design of the Driver's Workspace, *Am. J. Phys. Anthropol.,* vol. 16, no. 1, pp. 1–24, 1958.

18. ——— and R. G. Domey: Biotechnical Aspects of Driver Safety and Comfort, *Trans. SAE,* vol. 66, pp. 630–648, 1958.

19. ——— and ———: Experimental Studies of Night Vision as a Function of Age and Changes in Illumination, *Highway Research Board Bull.,* vol. 191, pp. 17–32, 1958.

20. ——— and A. L. Moseley: Human Factors in Highway Transport Safety, Harvard School of Public Health, Boston, 1954.

21. ——— and H. W. Stoudt: Human Body Size and Passenger Vehicle Design, *SAE Spec. Publ.* 142A, 1961.

22. ——— and ———: Physical Variables Influencing Driver Comfort, Efficiency and Safety, Harvard School of Public Health, Boston, 1955.

23. ——— et al.: Dark Adaptation as a Function of Age: I. A Statistical Analysis, *J. Gerontol.*, vol. 15, no. 2, pp. 149–154, 1960.

24. ——— et al.: Human Factors in the Design of Highway Transport Equipment, Harvard School of Public Health, Boston, 1953.

25. Norman, L. G.: One Hundred Thousand Deaths on the Roads of the World Every Year, *World Health*, vol. 15, no. 1, pp. 24–29, 1962.

26. Sendors, V. L., and J. Cohen: The Influence of Methodology on Research on Instrument Displays, *USAF WADC Tech. Rept.* 53-93, April, 1953.

27. Sinaiko, H. W., and E. P. Buckley: Human Factors in the Design of Systems, *Naval Research Lab. Rept.* 4996, 1957.

28. Summary Report of Automotive Crash Injury Research at Cornell University, U.S. Government, Commission on Accidental Trauma, Armed Forces Epidemiological Board, Washington, 1961.

29. Sutro, P. J., H. O. Ward, and C. A. Townsend: Human Visual Capacities as a Basis for the Safer Design of Vehicles, Commission on Accidental Trauma, Armed Forces Epidemiological Board, Washington, 1958.

30. U.S. Public Health Service: Accidental Injury Statistics, Selected Data, May, 1960, U.S. Department of Health, Education and Welfare, Washington, 1960.

Chapter 13

SELECTED PAPERS FROM THE AUTO CRASH INJURY RESEARCH OF CORNELL UNIVERSITY

The Staff of ACIR, B. J. Campbell, Editor*

THE DATA USED BY AUTOMOTIVE CRASH INJURY RESEARCH OF CORNELL University (ACIR) come from state police organizations that investigate accidents and from physicians who treat the injured. In states participating in the Cornell program, troopers fill out special ACIR forms when investigating certain accidents. These supplemental forms are necessary because standard state forms, having a different purpose, do not usually include the information necessary for ACIR research. In addition to completing the forms, troopers also photograph damage to the interior and exterior of the car. Physicians who care for the injured fill out special ACIR forms that detail the nature and severity of injuries. When these forms reach ACIR, they are analyzed and information is transferred to punched cards. Tabulations from these cards form the basis of research efforts.

To date, 24 states and 5 cities have contributed data, with the result that information is now available concerning 50,000 accident-involved cars (27,000 of which had one or more injuries) and more than 75,000 occupants. When a state enters the ACIR program, a sampling plan is established in order to obtain data that properly represent the state. Sometimes areas are selected randomly and shifted periodically until the state is reasonably well covered. In other cases, it is not possible to select areas with complete freedom, but efforts are made to select areas in such a manner as to obtain a reasonably representative sample.

Three groups of ACIR studies are particularly enlightening: evalua-

* Automotive Crash Injury Research of the Cornell Aeronautical Laboratory, Buffalo, N.Y. Formerly this organization was called Automotive Crash Injury Research of Cornell University and was located in New York, N.Y.

This chapter is derived from two ACIR reports: "A Summary of Selected Papers from Automotive Crash Injury Research of Cornell University," January, 1961, and "Summary Report, Automotive Crash Injury Research of Cornell University, 1953–1961," Apr. 3, 1961.

tions of ACIR's methodology, passenger-retention studies, and analyses of accident and injury factors.

EVALUATION STUDY

A Comparison of Automotive Crash Injury Research Samples with Complete State Data [2]

The distributions of ACIR data and data from certain states with respect to certain variables common to both were compared, as shown in Table 13-1. Thirty-one chi-square comparisons were made between

TABLE 13-1 *Comparison of ACIR with State Data*

Accident factor	Type of accident compared	States in which comparison was available			
		Ariz.	Calif.	N.C.	Tex.
Day of week	Fatal	X	X	X	X
Hour of day	Fatal	X	X	X	X
Month of year	Injury-producing*	. . .	X†	X†	X
Number of vehicles involved	Injury-producing	X	X		
Traveling speed	Fatal	. . .	X†	X	
Age of occupants	Injury-producing	X	X	X	X
Sex of occupants	Injury-producing	X	X	X	X
Age of vehicle	Fatal	. . .	. . .	. . .	X
Year of manufacture	Fatal	X			
Road surface	Injury-producing	X†			
Road character	Injury-producing	X			
Proportion of fatalities among all injury-producing accidents	. . . .	X	X	X	X

* Injury-producing tabulations include fatalities.

† Significant difference between ACIR and state data.

SOURCE: B. J. Campbell, "A Comparison of Automotive Crash Injury Research Samples with Complete State Data," Automotive Crash Injury Research of Cornell University, February, 1961.

the ACIR and the state data. Significant differences were found in four cases: (1) between ACIR and California data concerning the number of injury-producing accidents by month and (2) the number of fatal accidents by traveling speed; (3) between ACIR and North Carolina data concerning the number of injury-producing accidents by month; (4) between ACIR and Arizona data for injury-producing accidents by

type of road surface. Two of these differences, both concerning the distribution of accidents by month, merely reflect ACIR's practice of shifting sampling areas from one section of the state to another and do not indicate sampling defects. In the other two cases no clear explanation is apparent. However, 27 of the 31 comparisons indicated similarity, and it was concluded that ACIR data adequately represented accident events in these four states and presumably in the others.

PASSENGER-RETENTION STUDIES

Hugh De Haven first enunciated the principles of passenger "packaging" as a means of illustrating desirable crash-safety characteristics of automobiles [3]. He listed four principles that underlie the packaging of a valuable object for shipment under expected conditions of rough handling. First, the package should be constructed so as to resist crushing. Second, the outer shell of the package should not break open and spill the object. Third, the object should be securely fastened to the package so that, even if the shell does burst, the object will still be retained. Fourth, the package should be lined with energy-absorbing material so that, if the object moves within the package, it will not be damaged.

ACIR's passenger-retention studies began with the discovery in 1954 of a major problem area—the relation between door opening, occupant ejection, and serious injury. The basic relations that emerged in the following study have remained stable since that time.

A Study of Automobile Doors Opening under Crash Conditions [6]

An analysis was made of the relation between automobile door opening, occupant ejection, and degree of injury. Two samples, totaling 337 cases (a case consisted of a car with its occupants), were examined.[1] One sample consisted of 177 cars involved in injury-producing accidents in the states of North Carolina and Maryland. The second sample consisted of 160 cases involved in fatal accidents in the state of Indiana.

Examination of the data revealed that one or both front doors opened in 72 per cent of the fatal-accident sample and in 44 per cent of the injury-producing sample. When a rollover was the principal impact (18 per cent of the 337 cases), 77 per cent of the cars had at least one front-door opening, as compared with 53 per cent for nonrollovers. Right-corner impacts caused right front doors to open nearly twice as

[1] Although data volume was low, the findings were confirmed consistently until design changes were incorporated in automobiles.

often as left front doors; left-corner impacts caused left front doors to open nearly twice as often as right front doors. When the compartment area was hit, the door opposite the impact site opened in 49 per cent of the cases. As might be expected, the higher the impact speed or accident severity, the higher the relative frequency of door openings.

This study also pointed out the consequences of front-door openings —25 per cent of the occupants in cars where a front door opened were ejected in the injury-producing sample and 41 per cent in the fatal sample. Ejectees fared considerably worse than nonejectees. In the injury-producing sample, 60 per cent of the ejectees and only 26 per cent of the nonejectees suffered dangerous or fatal injuries; in the fatal sample, 91 per cent of ejectees experienced dangerous or fatal injury as compared with 76 per cent for the nonejectees.

The implications of the study seem clear: if doors could be kept closed, there would be a sizable—although as yet unestimated—reduction in the injury consequences of automobile accidents.

In 1956 all American automobile manufacturers introduced modified door locks designed to reduce the frequency of door opening in accidents. Preliminary examination of available ACIR data in 1957 indicated that, on an industry-wide basis, these locks appeared to be effective: the frequency of one or more front-door openings among the later models was about 27 per cent less than among older cars.

The next logical step was verification of the danger of ejection and an estimation of the potential saving of lives if ejection were prevented.

Ejection and Automobile Fatalities [8]

The study of automobile door openings as related to crash conditions had indicated the importance of occupant ejection in the injury picture. By using data from a sample of 3,261 pre-1956 automobiles and 7,337 occupants, a more detailed examination of the ejection-injury phenomenon was undertaken. The study proposed to deal with the fallacy that being "thrown clear" of a car during an accident would generally save one's life. The study attempted to answer the following questions:

1. Is the risk of fatal injury greater for occupants ejected from automobiles in an accident than for those who remain inside?

2. What would be the expected frequency of fatality for ejectees if they had remained inside the cars?

3. If ejection increases the risk of fatality, what is the estimated number of lives that could be saved annually by preventing ejection in injury-producing accidents?

This study confirmed the earlier finding of Moore and Tourin by giving an affirmative answer to the first question—the frequency of fatal

injuries among ejectees (12.1 per cent) was significantly higher than among nonejectees (2.5 per cent).

Since ejection occurred more frequently in severe accidents and since the risk of ejection varied according to occupant seating position, the comparison of ejectees and nonejectees was adjusted to establish comparable conditions of accident severity and seating position. For an estimation of the expected frequency of fatality if no ejection occurred, the risk of fatality for nonejected occupants was applied to occupants ejected under comparable conditions. That is, it was assumed that, had the ejectees remained in the car, they would have been exposed to this nonejected risk. The observed and expected fatalities for each seating position and degree of accident severity were each totaled, and it was calculated that the difference between the two—68 fatalities—represented the lives that would have been saved in the 3,261 accidents examined if ejection through open doors had been completely prevented.

This reduction—about 25 per cent—was then applied to national fatality figures for passenger-car occupants in order to estimate the saving in human lives if ejection were completely controlled. After suitable adjustments for the rural composition of the sample and for other variables, it was estimated that approximately 5,500 lives could be saved annually through the prevention of occupant ejection.

The emphasis thus far has been placed on door opening as a causative factor in injury to ejectees. There are at least two possible methods of reducing such ejection—the use of seat belts and the use of improved door locks. The next studies examined the effectiveness of seat belts in rural California accidents.

An unusual opportunity was afforded when the California Highway Patrol (CHP) was directed by the California Legislature to perform a study of seat-belt effectiveness in accidents investigated by the patrol during 1958. The patrol investigates about 40 per cent of the total accidents occurring in the state of California which account for 60 per cent of the fatal accidents in the state. A cooperative effort by ACIR and the California group resulted in the completion of two studies.

A Report on Safety Belts to the California Legislature: Summary and Analysis of California Highway Patrol Reports and Opinions on 54,348 Automobile Accidents [9]

This paper examined the availability and use of seat belts. The findings of the report were as follows:

1. Approximately 3.5 per cent of automobiles involved in accidents investigated by the patrol contained one or more safety belts. The avail-

ability of this safety device varied according to the seating areas in the automobiles. About 3.5 per cent of drivers and right-front-seat occupants had this device; less than 2.0 per cent of center-front- and rear-seat occupants had belts available.

2. Among cars equipped with safety belts, only one-third had all the installed belts worn by their occupants. Belt utilization differed among the occupants according to the various seating areas, with the drivers using their belts about one-third of the time, the right-front-seat occupants slightly less often, and the occupants of other seating areas using available belts less than one in six times.

Safety Belt Effectiveness in Rural California Automobile Accidents: A Comparison of Injuries to Users and Non-Users of Safety Belts [10]

Using data collected by the patrol, ACIR compared injuries to 933 persons who had worn seat belts and 8,784 persons who had not worn seat belts during an accident. In view of the possibility that the seat-belt group was somehow not comparable with the nonbelt group, several checks were made. It was found that average speed prior to an accident was somewhat higher for the belt group (a small but statistically significant margin of about 4 mph). Moreover, the seat-belt group was involved in significantly more rollover accidents, a type accompanied by marked risk of ejection (with an attendant increased risk of serious or fatal injury). Thus, the seat-belt cases, as a group, could be said to be composed of somewhat more serious accidents than the nonbelt group.

Since the two groups differed somewhat in terms of accident seriousness, it was necessary to analyze the accidents under conditions of greater comparability. Accordingly, a comparison was made of the frequencies of major-fatal injuries under various specific accident circumstances. This comparison and the resulting chi-square analysis are shown in Table 13-2. The incidence of major-fatal injuries was significantly less among seat-belt users ($P = < 0.01$). Regrouping the figures will show that the greatest difference between the groups was for rollover and nonspecific (lateral impacts, angled collisions, etc.) accidents. The data may also be regrouped to show speed and seat-position effects.

The study showed that the use of seat belts reduced the incidence of major-fatal injuries by about 35 per cent—largely through the prevention of ejection. The study also indicated that belt wearers were injured as often as nonbelt wearers but that the degree of injury was lower for belt wearers.

An evaluation of the second possible method for reducing occupant ejection, the use of improved door locks, was reported in 1961.

TABLE 13-2 *Observed and Expected "Major-Fatal" Injuries in Each Seat-belt Group*

Seat	Travel speed range, mph	Accident type	Non-seat-belt group: Number persons	Non-seat-belt group: Frequency major-fatal injury	Seat-belt group: Number persons	Seat-belt group: Frequency major-fatal injury	Seat-belt group: Expected major-fatal injury
Driver.........	0–30	Front-front	370	6	34	0	0.55
Driver.........	0–30	Front-rear	657	1	79	0	0.12
Driver.........	0–30	Rollover	167	10	14	1	0.84
Driver.........	0–30	Other	1,123	6	81	0	0.43
Driver.........	31–60	Front-front	540	40	69	5	5.11
Driver.........	31–60	Front-rear	1,027	25	125	4	3.04
Driver.........	31–60	Rollover	792	83	105	5	11.00
Driver.........	31–60	Other	1,184	60	116	2	5.88
Driver.........	61+	Front-front	38	13	7	2	2.39
Driver.........	61+	Front-rear	87	8	21	1	1.93
Driver.........	61+	Rollover	162	39	26	4	6.26
Driver.........	61+	Other	36	6	13	1	2.17
Right front....	0–30	Front-front	125	3	16	1	0.38
Right front....	0–30	Front-rear	276	1	29	0	0.10
Right front....	0–30	Rollover	60	5	4	0	0.33
Right front....	0–30	Other	397	12	30	0	0.91
Right front....	31–60	Front-front	240	24	30	6	3.00
Right front....	31–60	Front-rear	498	17	49	2	1.67
Right front....	31–60	Rollover	340	49	28	1	4.04
Right front....	31–60	Other	510	50	30	2	2.94
Right front....	61+	Front-front	16	6	4	0	1.50
Right front....	61+	Front-rear	50	4	7	0	0.56
Right front....	61+	Rollover	74	22	10	1	2.97
Right front....	61+	Other	15	1	6	0	0.40
Total........	.....	....	8,784	491	933	38	58.52

$$\chi^2 = \frac{(\text{observed} - \text{expected})^2}{\text{expected}} = \frac{(38.00 - 58.52)^2}{58.52} = 7.195, \qquad P = <0.01.$$

SOURCE: Boris Tourin and John W. Garrett, "Safety Belt Effectiveness in Rural California Automobile Accidents: A Comparison of Injuries to Users and Non-users of Safety Belts," Automotive Crash Injury Research of Cornell University, February, 1960.

An Evaluation of Door Lock Effectiveness: Pre-1956 vs. Post-1955 Automobiles [4]

This study of automobile doors compared post-1955 American automobiles, which had modified door locks, with older-model cars in order to determine whether the more recent models demonstrated a reduction in door opening, occupant ejection, and dangerous or fatal injuries.

The data examined in this study were drawn from the ACIR interstate program of rural injury-producing accidents and consisted of 14,135 American-made automobiles, in each of which at least 1 occupant had been injured. These cars had carried a total of 31,855 occupants. For purposes of comparison, the data were divided according to year of manufacture—pre-1956 cars (8,606) and post-1955 cars (5,529).

The two groups of cars were then compared under similar conditions of highest impact speed (in the accident), principal area of impact, and body style in order to determine whether the frequency of one or more front-door openings, occupant ejection, and dangerous or fatal injury had been decreased by the use of modified door locks. This comparison is

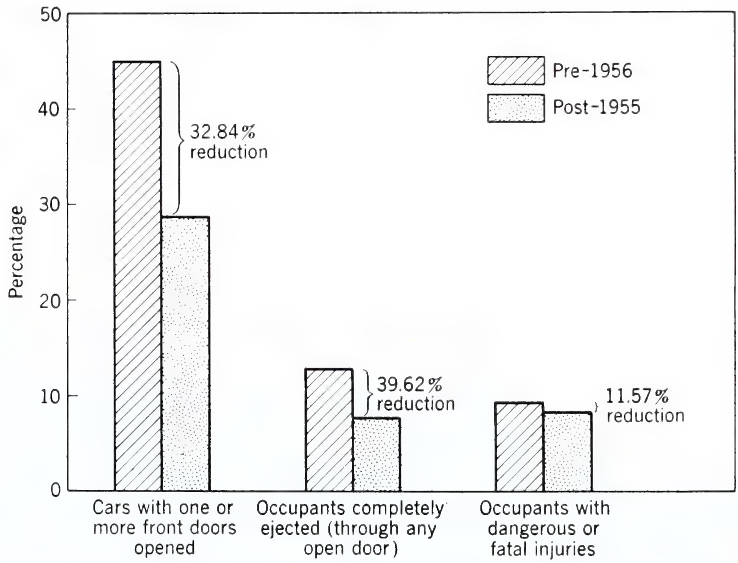


FIG. 13-1 *Frequency of door opening, occupant ejection, and dangerous or fatal injury (pre-1956 vs. post-1955 automobiles). (After John W. Garrett, An Evaluation of Door Lock Effectiveness: Pre-1956 vs. Post-1955 Automobiles, Automotive Crash Injury Research of Cornell University, July, 1961.)*

shown in Figure 13-1. It was estimated that among the later-model cars:

1. The frequency of door opening was reduced about 33 per cent.

2. The frequency of occupant ejection was reduced about 40 per cent.

3. The frequency of dangerous and fatal injuries was reduced about 12 per cent.

4. Complete prevention of ejection—by means of a "perfect" door lock—would reduce fatality by approximately 25 per cent. Based on a combination of dangerous and fatal injuries, this finding confirmed the estimate made in Tourin's 1958 study [8].

5. The modified door locks in later-model cars achieved about 35 per cent of the potential injury-reduction possible if no door opened.

Based on the earlier estimate of saving 5,500 lives via "perfect" locks installed in all cars, it can be roughly calculated that if all cars on the highways had been post-1955 models—with "modified" door locks—an annual saving of approximately 1,900 lives might result.

STUDIES OF ACCIDENT AND INJURY FACTORS

Although of necessity nearly all ACIR reports are concerned with occupant injury, one class of studies may be regarded as medically oriented. These studies vary from broad presentations of injury data—specific types of injury and injury severity, body area injured, etc.—to specific studies of injury types and the relation between occupant characteristics and injury. During the early days of its program, ACIR developed

TABLE 13-3 *ACIR Injury Scale*

Degree	Type of Injury
Noninjured	
Minor	Contusions, abrasions, superficial lacerations
Moderate	Deep lacerations, simple bone fractures, joint or muscle sprain
Severe	Extensive lacerations, compound fractures with bone displacement
Serious	Lacerations with dangerous hemorrhage, crushing and compression fractures
Critical	Dangerous chest or abdominal injury, spinal-cord damage
Fatal No. 1	Fatal lesions to one body area, with severe or lesser injuries elsewhere
Fatal No. 2	Fatal lesions to one body area and serious or critical injuries elsewhere
Fatal No. 3	Fatal lesions to two body areas
Fatal No. 4	Fatal lesions to three or more body areas

SOURCE: "Summary Report," Automotive Crash Injury Research of Cornell University, 1953–1961, Apr. 3, 1961.

two qualitative scales for classifying injury and accident severity. These are shown in Tables 13-3 and 13-4.

TABLE 13-4 *ACIR Accident-severity Scale*

Degree	Type of accident damage
Minor	Dents or deformation of car body, with basic structure undamaged
Moderate	Buckling or crumpling of car body, with basic structure undamaged
Moderately severe	Crushing of car body, with distortion of basic structure
Severe	Much damage to car body and structure, some invasion of compartment area
Extremely severe	Partial demolition, with serious compartment-area invasion
Extreme	Complete demolition, disintegration of body and structure

SOURCE: "Summary Report," Automotive Crash Injury Research of Cornell University, 1953–1961, Apr. 3, 1961.

The listed definitions of severity are quite gross. The actual analytical criteria are extremely complex and require intensive training in the use of data and case photographs.

NOTE: The last three classes of damage are further subclassified into four degrees of survivability that are based on what chance an occupant would have had to survive in the same accident situation.

Medical Aspects of Automotive Crash Injury Research [1]

This study concerns injuries associated with automobile accidents. For the purpose of providing some indication of medical findings that might be expected in the 1 million persons injured in automobile accidents yearly, an examination was made of medical data from a sample of 1,000 injury-producing accidents. These accidents, occurring over the period 1952 to 1955, involved 2,253 occupants, of whom 1,678 (74.5 per cent) were injured to some degree.

Injury was first analyzed according to the overall description of injury shown in Table 13-5. About 50 per cent of those injured in the sample suffered minor injuries requiring little medical attention. On the other hand, 6.5 per cent sustained fatal injury, and another 6.4 per cent received injuries classified as serious or critical. At least one-third of the injured required medical care, with probably half these demanding hospitalization. Approximately two-thirds of the injured suffered injury in two or more of the six major body areas, indicating that the chance of multiple injuries was more than 50 per cent. Head injury was by far the most frequent in this sample. The frequency and degree of injury to the

TABLE 13-5 *Per Cent Distribution of Overall Degree of Injury Sustained by 1,678 Persons*

Degree of injury	Percentage
Minor	47.7
Nondangerous:	
Moderate	19.4
Severe	6.6
Dangerous:	
Serious	3.5
Critical	2.9
Fatal	6.5
Survived, injury unknown	13.4
	100.0

SOURCE: Paul W. Braunstein, Medical Aspects of Automotive Crash Injury Research, *J. Am. Med. Assoc.*, vol. 163, no. 4, pp. 249–255, January, 1957.

six body areas are shown in Table 13-6. As seen in this table, 72.3 per cent of the injured persons received head injuries, of which 61.8 per cent were in the minor-injury category. Although only 3.0 per cent and 3.9 per cent, respectively, sustained dangerous and fatal injuries to the head, on the basis of 1 million automobile accidents a year this would amount to approximately 69,000 such persons, many of whom would require specialized neurosurgical care. Moreover, disfigurements caused by the soft-tissue and facial-bone injuries suffered by many of these persons would present serious social and psychological problems to the injured.

Prominent among the other affected body areas were the thorax and thoracic spine, accounting for 36.6 per cent of the injuries and 2.0 per cent of the fatalities. High among the thoracic injuries were the "crushed-chest" type of injury (10 per cent), which was probably responsible for many of the fatalities in this group. Only 6.8 per cent of the injured incurred neck and cervical spine injuries, but 16.7 per cent of these injuries proved fatal—by far the highest rate for lethal injury. On the other end of the scale were skeletal injuries in the upper extremities, which, while accounting for the same percentage of injured, showed hardly any incidence of dangerous or fatal injury.

Injury to lower extremities was the second most frequent, occurring in almost half the cases and usually in conjunction with head injury. As in the case of the upper extremities, the incidence of dangerous and fatal injury was found to be relatively rare. Not so the experience of injuries to the abdomen and pelvis, which occurred only 15.3 per cent of the time but manifested a high incidence of dangerous to fatal injury (16.7 per cent), second only to those affecting the neck and cervical spine area. The latter two areas have also been the center of much interest in regard

TABLE 13-6 *Relative Frequency of Injury to Six Body Areas and Per Cent Distribution of Degree of Injury*

Degree of injury	Area					
	Head	Neck	Thorax	Abdomen and pelvis	Upper extremities	Lower extremities
Minor	44.7 (61.8)*	3.0 (43.9)	19.2 (52.6)	7.1 (46.3)	20.5 (69.6)	34.3 (73.0)
Nondangerous	12.9 (17.9)	1.0 (14.9)	8.1 (22.2)	3.5 (22.6)	5.9 (20.0)	8.2 (17.5)
Dangerous	3.0 (4.2)	0.4 (6.1)	2.2 (6.0)	2.0 (13.2)	0.1 (0.4)	0.2 (0.4)
Fatal	3.9 (5.4)	1.1 (16.7)	2.0 (5.4)	0.5 (3.5)	0.0 (0.0)	0.2 (0.4)
Not reported	7.8 (10.8)	1.3 (18.4)	5.1 (13.8)	2.2 (14.4)	2.9 (9.9)	4.1 (8.7)
Total	72.3 (100.1)	6.8 (100.0)	36.6 (100.0)	15.3 (100.0)	29.4 (99.9)	47.0 (100.0)

* Numbers outside the parentheses refer to the proportion of occupants having a particular type of injury. Thus, 44.7 per cent of the people had minor head injuries. Since the average number of injuries per person was about 2, these numbers add to approximately 200 per cent. Numbers in parentheses show the distribution of injuries by body area. Thus, 44.7 per cent of the people had *minor* head injuries, and these constituted 61.8 per cent of all head injuries.

SOURCE: Paul W. Braunstein, Medical Aspects of Automotive Crash Injury Research, *J. Am. Med. Assoc.*, vol. 163, no. 4, pp. 249–255, January, 1957.

to what effect seat belts would have on the incidence of cervical whiplash and pelvic snubbing injuries.

Child Injuries in Automobile Accidents [7]

The data examined in this study were collected on rural highways and consisted of 14,520 automobiles involved in nonpedestrian accidents in which at least one car occupant had been injured. Of the 31,001 occupants of known age in these cars 75.7 per cent had been injured. For purposes of this study, occupants were classified according to age as follows:

	Per cent
Children (to 11 years)	8.8
Adolescents (12 to 18 years)	15.5
Adults (over 18 years)	72.8
Age unknown	2.9
Total	100.0

The purpose of the study was to examine the three age groups for possible differences in injury frequency, injury severity, or patterns of injury to specific body areas. The influence of seat position on injury for each age group was examined, and possible evidence that other factors influenced the observed differences was sought for future guidance.

The following observations were made:

1. The proportion of children in injury-producing accidents is markedly lower than the proportion of the national population that they represent.

2. Children in automobile accidents are injured less frequently and less severely than adolescents and adults. As "age" increases, the frequency and severity of injury also increase.

3. Head injuries are somewhat more frequent among children than among adolescents or adults; injuries to other body areas are less frequent among children.

4. Children's head injuries are milder than those of older persons; their lower-extremity injuries are more severe; injuries to upper extremities are roughly the same in severity for children and others; thoracic, abdominal, and cervical injuries of children are milder.

5. In each seating area of the car, injury increases with age.

6. In all age groups, injury associates in the same fashion with seat area occupied; rear seats are "safer" than all others, and front seats are "safer" than "unusual" seating positions (standing, lying, etc.).

7. It is probable that the comparatively mild injury experience of

children can be further reduced by exploiting the protective aspects of rear-seat positions.

Perhaps one of the most controversial automotive-safety subjects is the role of speed in the crash-injury picture. In order to provide some perspective on this subject, ACIR examined a number of the problems associated with speed and injury.

A Study of Speed in Injury-producing Accidents: A Preliminary Report [5]

The common assumption that fatal injuries sustained during automobile accidents are associated with high speed and that rigid control of speed could eliminate fatal injury was explored in this study. Data on 3,203 automobiles and 7,154 occupants involved in injury-producing accidents during 1953 to 1956 were examined to determine the effect of speed on dangerous or fatal injuries. The study consisted of four parts: (1) frequency distribution of cars according to "traveling" and "impact" speeds in 10-mph steps; (2) relative frequency of dangerous or fatal injury in each speed class; (3) extent to which control of speed would reduce the frequency of dangerous or fatal injuries; (4) influence of factors other than speed (ejection, seat area occupied, and site of impact).

Among the 7,154 occupants, the various degrees of injury were distributed as follows:

	Percentage
No injury	25
Minor injury (bruises, minor lacerations, etc.)	46
Nondangerous injury (severe lacerations, fracture, etc.)	20
Dangerous injury (internal injuries, brain injuries, etc.)	5
Fatal injury (death instantaneous or within 24 hr after accident)	4
Total	100

As Figure 13-2 shows, the greatest proportion of injury-producing cars had been traveling at speeds in the 50- to 59-mph range and had impacted in the 40- to 49-mph range. About 74 per cent of the cars studied had been moving at speeds under 60 mph before the accident. About 70.9 per cent of the cars had impacted at speeds below 50 mph.

Figure 13-3 shows the proportion of persons traveling or impacting at successive speed ranges who suffered dangerous or fatal injury. (Such injuries were sustained by 9 per cent of the total number of occupants in the cars studied.) These two grades of injury fall at the severe end of the injury range. Injury increases were relatively small in speed ranges up to and including 40 to 49 mph, but above this range the frequency became more marked. The relative frequency rose to 28 per cent among

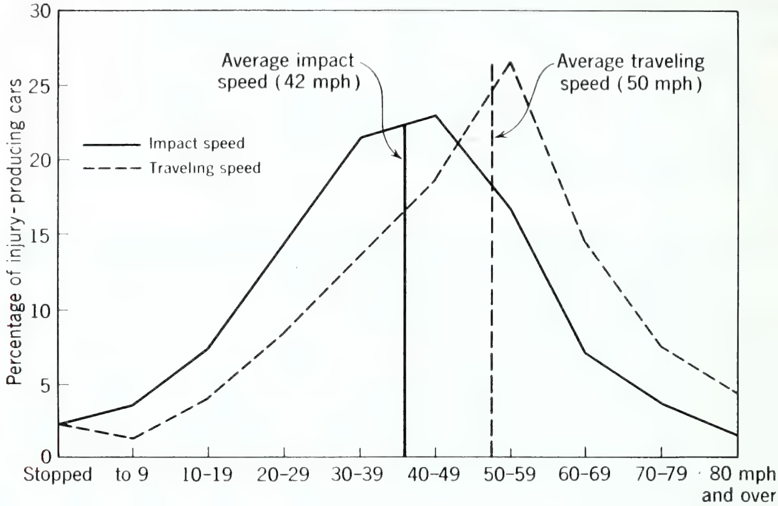


FIG. 13-2 *Distribution of reported traveling and impact speeds among 3,203 cars in injury-producing accidents. (After John O. Moore, A Study of Speed in Injury-producing Accidents: A Preliminary Report, Am. J. Public Health, vol. 48, no. 11, pp. 1516–1525, November, 1958. By permission of the publishers.)*

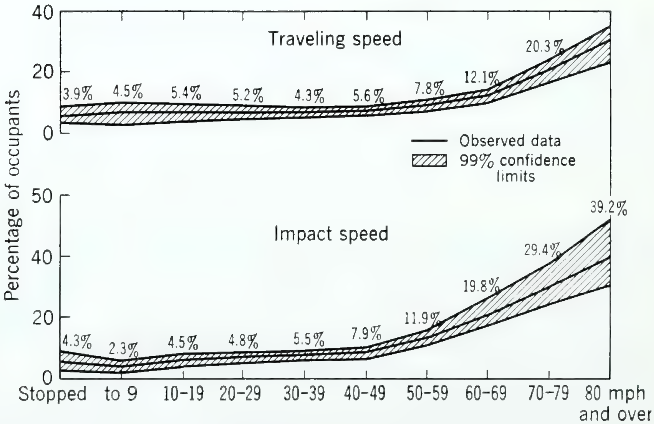


FIG. 13-3 *Frequency of dangerous or fatal injury according to progressively higher ranges of traveling and impact speeds. (After John O. Moore, A Study of Speed in Injury-producing Accidents: A Preliminary Report, Am. J. Public Health, vol. 48, no. 11, pp. 1516–1525, November, 1958. By permission of the publishers.)*

persons traveling at or above 80 mph and to 40 per cent among persons in cars impacting at this speed.

To estimate the injury effects of rigidly controlling speed, it was assumed that if cars traveling at high speeds were slowed down to some specified speed the proportion of dangerous-fatal injuries would be reduced to that observed at the specified speed. Thus, if speed were controlled so that no car traveled over 79 mph, 96 per cent of the observed dangerous-fatal injuries would still occur. It should be noted that reducing speed might prevent some of the accidents, but there is no way of estimating this from data in the present study. These calculations are based on the assumption that the accidents would occur. Thus, a car instead of crashing at a higher speed is assumed to crash at a lower speed.

Other extrapolations of speed and injury are listed as follows:

Speed, mph, limited to	Percentage of dangerous-fatal injuries remaining (100% equals that under existing conditions)
79	96
69	85
59	73
49	60

Examination of factors other than speed revealed the following. The risk of dangerous or fatal injury was greatest for the right-front-seat passenger (10.4 per cent of right-front occupants suffered dangerous or fatal injury), followed by the driver (7.5 per cent). (Only drivers with passengers in the car were considered.) Ejectees were five times as likely to suffer fatal injury as nonejectees. Passengers at or near the crash-impact site had a much higher risk of experiencing a dangerous-fatal injury than other occupants.

It was concluded that speed is but one (albeit an important one) of many factors that affect the degree of injury. Therefore, automotive safety requires a multidisciplinary approach in order to cope with the many complex problems encountered.

BIBLIOGRAPHY

1. Braunstein, Paul W.: Medical Aspects of Automotive Crash Injury Research, *J. Am. Med. Assoc.*, vol. 163, no. 4, pp. 249–255, January, 1957.

2. Campbell, B. J.: "A Comparison of Automotive Crash Injury Research Samples with Complete State Data," Automotive Crash Injury Research of Cornell University, February, 1961.

3. De Haven, H.: Accident Survival—Airplane and Passenger Automobile, Department of Public Health and Preventive Medicine, Crash Injury Research, Cornell University Medical College, 1952.

4. Garrett, John W.: "An Evaluation of Door Lock Effectiveness: Pre-1956 vs. Post-1955 Automobiles," Automotive Crash Injury Research of Cornell University, July, 1961.

5. Moore, J. O.: A Study of Speed in Injury-producing Accidents: A Preliminary Report, *Am. J. Public Health*, vol. 48, no. 11, pp. 1516–1525, November, 1958.

6. ——— and Boris Tourin: A Study of Automobile Doors Opening under Crash Conditions, Automotive Crash Injury Research, Cornell University Medical College, August, 1954.

7. ——— et al.: Child Injuries in Automobile Accidents, Automotive Crash Injury Research of Cornell University, February, 1960.

8. Tourin, Boris: Ejection and Automobile Fatalities, *Public Health Rept.*, vol. 73, no. 5, pp. 381–391, May, 1958.

9. ——— and John W. Garrett: A Report on Safety Belts to the California Legislature: Summary and Analysis of California Highway Patrol Reports and Opinions on 54,348 Automobile Accidents, Automotive Crash Injury Research of Cornell University, February, 1960.

10. ——— and ———: Safety Belt Effectiveness in Rural California Automobile Accidents: A Comparison of Injuries to Users and Non-users of Safety Belts, Automotive Crash Injury Research of Cornell University, February, 1960.

Chapter 14

THE OPERATOR AND VEHICLE DESIGN

Richard G. Domey and Ross A. McFarland**

THE PRIMARY FUNCTION OF HIGHWAY TRANSPORTATION IS THE RAPID movement of people and goods from one place to another. The efficiency of the system may be measured by (1) the degree of delay relative to the potential rate of traffic movement and (2) the frequency and severity of accidents resulting in injury, death, and property damage.

The achievement of transport objectives is dependent upon the interaction of four distinct units, (1) the human operator, (2) the vehicle, and (3) the highway complex, including (4) the meteorological environment. Inherent in each of these four elements are certain functions and responsibilities upon which rests the value of the operational system. Contrary to general opinion, these four units are not functionally independent.

Vehicle design is only one part of an extremely complex transportation system, and the geometry of operator workspace and passenger compartments composes only one facet of vehicle design. However, it is a particularly important facet. The incongruity between operator and machine must be minimized, for both precision of control and comfort are dependent upon compatibility of design and user.

HUMAN FACTORS IN VEHICLE DESIGN

The main hypothesis underlying the analysis and interpretation of the data to be presented is that maximum efficiency of machine operation depends upon the integration of man-machine-environment components into a cohesive unit or system, the separate parts or functions of which complement each other. The historical industrial-engineering and industrial-psychology data, reinforced by many extensive studies conducted by human factors specialists, have repeatedly shown that there is a close relation between design of operator workspace and efficiency of performance.

The design of seats, workspace layout, reach and effort requirements, and spatial locations must be carefully controlled if fatigue, monotony, and errors are to be reduced and comfort, safety, and efficiency are to

* Division of Environmental Health and Safety, Harvard School of Public Health, Boston, Mass.

be increased. Machines should be designed for the people who use them and for the environment in which both are expected to function. It follows that the efficient operation of machines, based on human perception and response, is dependent on the degree to which the machine has been designed to accommodate human physiological, structural, and functional factors. For instance, if seeing efficiency at night is dependent upon maximum acuity, then road users may sacrifice a margin of safety if they drive cars equipped with tinted glass that reduces luminance. Several recent technical papers presenting original experimental data on the effects of tinted windshields (filters) upon visual acuity, recovery from glare, stereopsis, and dark adaptation in groups varying widely in age have systematically demonstrated a decrement in seeing under reduced levels of luminance [2, 3, 5, 11, 17]. Yet tinted windshields that filter 30 per cent luminance are in common use, especially in domestic vehicles.

The present investigation, the eighth of a series,[1] is a human factors analysis of 20 domestic and 22 foreign passenger cars. This range permitted systematic comparisons among domestic cars and vehicles imported from England, France, West Germany, Sweden, and Italy.

In general, all vehicle manufacturers design machines to be sold to the operator population at large. Therefore, it is expected that vehicle designers have developed operator-workspace criteria congruent with the ranges of anthropometric differences characteristic of the consumer population. These criteria should not vary in any important dimension from one manufacturer to another, nor should they vary among the several sections and divisions within the individual companies.

The test of this expectancy depends on several conditions. First, it is necessary to describe the range of physical anthropometric differences known to exist in the vehicle-operator population and from these data to derive criteria useful in the evaluation of operator-workspace design. To a certain extent this has been done in the United States for operators of commercial as well as military vehicles. In combination, these operators compose a large portion of the general driver population.

[1] The findings of the first of seven vehicular studies were published in a report [13] in which detailed evaluations were made of 12 trucks and six buses (1950 and 1951 models). Also several preliminary design criteria were formulated. In "Human Body Size and Capabilities in the Design and Operation of Vehicular Equipment" [12] the basic approaches of human engineering and pertinent biological data were presented in some detail. The Application of Human Body Size Data to Vehicular Design [7] is a summary of much of this material and includes specific design recommendations for the driver workspace, as does "Human Factors in the Design of Vehicle Cab Areas" [4]. Also see Biotechnical Aspects of Driver Safety and Comfort [8]. A sixth study, entitled Human Factors in the Design of Trucks [14], is a modified and detailed follow-up analysis of vehicles manufactured in 1956. The seventh study, Human Factors in the Design of Passenger Cars: An Evaluation Study of Models Produced in 1957 [9], includes a human factors analysis of no less than 36 interior dimensions of 20 passenger cars.

Recently the known physical anthropometric studies of white Americans, male and female, have been collected and the various data summarized and integrated [16]. The U.S. Public Health Service in the National Health Survey is amassing certain limited anthropometric data (clothed subjects) from both males and females. However, it will be 1 or more years before the basic information will become generally available. Because of size differences between the North American and continental driver populations, standards derived from these data do not specifically apply to the European. However, American anthropometric human-sizing standards for vehicle workspace are applicable to foreign vehicles marketed in this country.

Second, it is necessary to obtain detailed measurements of the internal spatial dimensions of a fair sample of stock cars and to compare the obtained dimensions with available human-sizing criteria. This was done in 1957 [9] and was repeated in the study described in this paper.

At present there is little evidence to support the hypothesis that operator workspace in the several vehicles has been developed on the basis of systematic use of adequate biotechnical data and criteria common to the automotive industry or common to the several sections and divisions of individual companies. The data obtained in both the 1957 and the present studies varied from one vehicle to another, year by year, were often incongruent with human-sizing criteria, and were inconsistent both among and within the individual companies.

In a 1955 review of occupational anthropology, Damon and McFarland described in detail the physical characteristics of bus and truck drivers [1]. More recently, a tentative set of human-sizing criteria was developed from a sample of commercial-vehicle operators and a sample of military-vehicle operators [6]. In 1957 these criteria were used to evaluate the workspace design of many commercial vehicles and passenger cars. However, in the interim between 1957 and 1961, more detailed work was accomplished in describing the anthropometric characteristics of the American population. The data describing persons sixteen years of age and above are especially relevant to the problem of designing vehicular workspace. Two related papers are especially noteworthy: that by Stoudt, Damon, and McFarland [16], which combines and organizes the data from many anthropometric studies, and that by McFarland and Stoudt [10], which recommends sizing criteria for passenger vehicles.

HUMAN-SIZING CRITERIA

Table 14-1 shows the variation in size of the different body components of American males and females. The range from the 5th to the

95th percentile and the medians of 18 body dimensions are included. Criteria applicable to American drivers for workspace geometry in passenger cars were developed by McFarland and Stoudt [10] from the anthropometric data of Table 14-1. They are presented in Table 14-2. All criteria allow for the physical dimensions of the clothed driver.

TABLE 14-1 *Human Engineering Body Measurements of Passenger-car Drivers*

	Male drivers*			Female drivers*		
	Percentiles			Percentiles		
	5th	50th	95th	5th	50th	95th
1. Stature	64.1	68.4	72.6	59.5	63.4	67.3
2. Weight, lb	131	166	216	105	136	190
3. Sitting height	33.8	36.0	38.2	31.6	33.7	35.6
4. Eye height	29.3	31.6	33.7	27.2	29.3	31.1
5. Shoulder height	21.4	23.3	25.2	19.3	21.1	22.9
6. Elbow height	7.7	9.3	10.9	8.2	9.4	10.9
7. Thigh height†	4.8	5.7	6.8	4.9	5.8	6.7
8. Popliteal height	15.5	16.9	18.1	13.9	15.2	16.2
9. Knee height	19.8	21.6	23.5	17.9	19.5	20.8
10. Anterior arm reach	32.0	34.9	37.5	28.5	30.9	33.4
11. Abdomen depth†	8.4	10.1	12.4	7.9	9.0	11.1
12. Buttock-knee length	21.6	23.5	25.5	20.7	22.4	24.0
13. Buttock-popliteal length	17.4	18.9	20.8	16.8	18.2	20.0
14. Foot length	9.6	10.4	11.3	8.8	9.6	10.2
15. Shoulder breadth	16.4	17.7	19.4	14.3	15.6	17.5
16. Elbow breadth	15.0	17.4	20.6	13.4	15.0	16.8
17. Seat breadth	13.0	14.4	16.2	13.6	15.1	17.2
18. Foot breadth	3.6	3.9	4.3	3.3	3.6	3.9

* All data are in inches unless otherwise specified.

† Tentative data.

SOURCE: R. A. McFarland and H. W. Stoudt, Human Body Size and Passenger Vehicle Design, *SAE Spec. Publ.* 142A, January, 1961.

TABLE 14-2 *Passenger-vehicle Criteria for Driver's Seat and Workspace*

1. Seat height (vertical distance from floor to top of seat cushion):

Relevant body measurement. Popliteal height (vertical distance from floor to underside of thigh immediately behind knee, body erect, knees at right angle).

Recommendation. A maximum height of 14 in., with an optimum range of 10–14 in.

Discussion. Excessively high seats cause an uncomfortable and sometimes painful pressure on the under part of the thighs. Seat height must always be less than the popliteal height, including shoes, of the smallest drivers. Sixteen inches would be an acceptable maximum height if the feet were placed on the floor directly in front of the seat. However, in driving, the legs are extended forward, and popliteal height is effectively reduced; thus a lower seat height is required. Low seat heights, i.e., those under 10 in., are judged uncomfortable

TABLE 14-2 *Passenger-vehicle Criteria for Driver's Seat and Workspace (Continued)*

by some drivers. With vertically adjustable seats, the upper limit of the range should be set at 14 in., since it is the shorter drivers who tend to use the higher adjustments to increase their field of vision from the vehicle.

2. Seat length (fore-and-aft distance on surface of seat from most forward point on front of cushion to intersection of back rest):

Relevant body measurement. Buttock-popliteal length (horizontal distance from most posterior point on buttocks to back of lower leg at knee, body erect, knees at right angle).

Recommendation. 18 in.

Discussion. Seats that are excessively long cut into the back of the lower leg at the knee and restrict the range of knee movement. A length of 18 in. is the longest that can be comfortably occupied by smaller drivers, though even this length will force them to shift their buttocks forward on the seat almost 1 in. However, shortening the seat much below 18 in. will not provide adequate thigh support for larger drivers.

3. Seat breadth (lateral distance on seat surface between sides of seat):

Relevant body measurement. Hip breadth, seated (horizontal distance between most lateral points on buttocks).

Recommendations. (1) For an individual seat, a minimum of 18 in. (2) For side-by-side accommodation of three persons, optimally a breadth of 58 in.

Discussion. Additional space should always be provided for the individual seat, if possible, to permit greater freedom of movement and change of body position.

4. Back-rest height (distance of surface of back rest between seat cushion and top of back rest):

Relevant body measurement. Shoulder height (vertical distance between seat surface and top of shoulder, body erect).

Recommendation. 18–21 in.

Discussion. Back-rest height is not critical as long as the following two criteria are complied with: (1) The minimum height must be at least 12 in., in order to support the small of the back, and should be somewhat higher to provide additional support for the upper back. (2) The back rest must not be so high as to interfere with the driver's headgear or provide him with a headrest, since the latter is detrimental to both alertness and freedom of head movement. A height not in excess of 21 in. will fulfill these conditions. For passengers, of course, a headrest may be desirable.

5. Back-rest breadth (lateral distance on surface of back rest between sides of back rest):

Relevant body measurement. Shoulder breadth (maximum horizontal distance between most lateral points on shoulders).

Recommendations. (1) For an individual seat a minimum of 20 in. (2) For side-by-side accommodation of three persons, optimally a breadth of 58 in., though 54 in. may be acceptable if the former figure is excessive in relation to basic vehicle dimensions.

Discussion. Additional space should be provided, if possible, to permit greater freedom of movement and changes of body position.

TABLE 14-2 *Passenger-vehicle Criteria for Driver's Seat and Workspace (Continued)*

6. Seat-roof distance (vertical distance between lowest point on occupied seat surface and roof):

Relevant body measurement. Sitting height (vertical distance from seat surface to top of head, body erect).

Recommendation. 40 in.

Discussion. A minimum dimension: always provide additional clearance, if possible. It should be noted that this distance is measured from the bottom of the deflected seat cushion. A measurement from the top of the undeflected cushion is meaningless, unless the precise amount of cushion deflection (which can vary widely) is known. The above figure will provide almost 2 in. of clearance above the heads of the taller drivers when sitting erect, enough to accommodate most forms of headgear. If the seat is adjustable vertically, the 40-in. clearance can be from the point of lowest adjustment, since shorter drivers tend to use the higher positions to obtain maximum vision from the vehicle.

7. Seat–steering-wheel distance (vertical distance between bottom of steering wheel to occupied seat surface):

Relevant body measurement. Thigh thickness (vertical distance between seat surface and highest point on top of thigh).

Recommendation. 7 in.

Discussion. A minimum dimension: provide more clearance, if possible, but not at the expense of raising the wheel to an uncomfortably high position. It should be noted that this distance is measured to the deflected seat cushion. A measurement to the undeflected cushion is of little value, unless the amount of normal deflection is known. Inadequate clearance in this dimension will not permit larger drivers, when entering or leaving the car, to slide their legs under the steering wheel without discomfort.

8. Back-rest–steering-wheel distance (horizontal distance between back rest and nearest point on steering wheel):

Relevant body measurement. Abdomen depth (horizontal distance from back rest to most anterior point on abdomen, body seated but relaxed).

Recommendation. 14 in.

Discussion. With fore-and-aft seat adjustment, the 14-in. clearance should be measured from the middle of the range. Inadequate clearance will result in the steering wheel's indenting and rubbing the front of the abdomen.

9. Back-rest–dashboard distance (distance between back rest and dashboard or nearest forward obstruction at knee level in vertical fore-and-aft plane of knee):

Relevant body measurement. Buttock-knee length (horizontal distance between most posterior point on buttocks to most anterior point on knee, body erect, knees at right angle).

Recommendation. 27½ in.

Discussion. A minimum clearance: always provide additional space if possible. With fore-and-aft seat adjustability, this minimum distance can be measured from the rearmost adjustment used by taller, longer-legged drivers.

10. Pedal–steering-wheel (or shift-lever) distance (distance from pedal surface to nearest point on bottom of steering wheel or shift lever in vertical fore-and-aft plane of the pedal):

TABLE 14-2 *Passenger-vehicle Criteria for Driver's Seat and Workspace (Continued)*

Relevant body measurement. Knee height (vertical distance from floor to top of knee, knee and ankle at right angle).

Recommendation. 24 in.

Discussion. A minimum dimension: provide more space if possible, but not at the expense of raising the wheel or shift lever to an uncomfortably high position. Taller drivers, when shifting the foot from the accelerator to the brake pedal, often find their leg movements restricted because of inadequate leg room and are sometimes forced to angle the leg out to the right of the wheel and straighten the leg as the pedal is depressed. The significance of such a loss of time during a critical braking operation is obvious. It is important to note that this dimension should be measured to the nearest point on the wheel or shift lever immediately above the brake pedal.

11. Vertical seat adjustability:

Relevant body measurement. Eye height, body seated (vertical distance between seat surface and eye level, body in normal, but not necessarily erect, seated position).

Recommendation. 4½ in.

Discussion. Though often neglected, this adjustment is extremely important for placing the eye level of both tall and short drivers at the optimum vertical location for vision from the vehicle. The present frequent use of pillows by shorter drivers and the often hunched position of taller drivers attest to the need for such a seat adjustment. A vertical adjustment of 4½ in. matches the range of normal eye height of taller and shorter seated drivers. The adjustments should be made in increments of ½ in. or less. (See Seat height.)

12. Fore-and-aft seat adjustability:

Relevant body measurement. Functional leg reach (distance from most posterior point on buttocks to surface of the outstretched foot, body seated and erect). This measurement varies widely with knee angle.

Recommendation. 8 in.

Discussion. Fore-and-aft adjustment of the seat is essential so that long- and short-legged drivers are able to operate the vehicle's controls equally well. Functional leg length or reach is one of the most variable of human-body measurements and thus requires a wide range of seat adjustment. The adjustments should be made in increments of ½ in. or less.

13. Seat-surface angle:

Recommendation. 7° (from horizontal).

Discussion. Some rearward inclination of the seat surface is desirable to permit the back rest to support part of the weight of the body and to prevent the buttocks from sliding forward on the seat. The above angle will probably be found comfortable by most persons.

14. Back-rest angle:

Recommendation. 112° (from horizontal).

Discussion. This angulation of the back rest seems to approximate the general preference of most people, though adjustability of the angle by ±5° or more would provide considerable additional comfort by permitting changes of body position.

SOURCE: R. A. McFarland and H. W. Stoudt, Human Body Size and Passenger Vehicle Design, *SAE Spec. Publ.* 142-A, January, 1961.

TABLE 14-3 *Comparison of Selected Dimensions of the Driver's Workspace in American and Imported Automobiles (1959 Models)*

	Domestic vehicles (N = 20)*			Imported vehicles (N = 22)*		
	Maximum	Minimum	Range	Maximum	Minimum	Range
Windshield						
1. Windshield areas	1,888 in.²	894 in.²	994 in.²	1,308 in.²	444 in.²	864 in.²
2. Length of wiper blade	16	12	4	12¼	8	4¼
3. Area of windshield cleaned by one blade	475.5 in.²	167.2 in.²	308.3 in.²	392.5 in.²	142.6 in.²	249.9 in.²
4. Percentage of area cleaned by one wiper blade	25.7%	15.3%	10.4%	51.2%	10.9%	40.3%
Seat Dimensions						
5. Breadth of front seat:						
Full	60	55¼	4¾	50¼	44½	5¾
Split	22	21¾	¼	27¼	18	9¼
6. Depth of front seat	19¾	17	2¾	21	17	4
7. Height of front seat above floor	12¾	7½	5¼	14⅝	8	6⅝
8. Breadth of front-seat back:						
Full	55½	53¼	2¼	47¾	43¾	4
Split	27½	21	6½	25⅞	18¾	7⅛
9. Height of front-seat back	22	17¼	4¾	22¼	17	5¼
10. Angle of seat	100°	91°	9°	100°	87°	13°
11. Breadth of rear seat:						
Full	66¼	52	14¼	57½	34¼	23¼
Split	20	20	0			
12. Depth of rear seat	20	16	4	20	15⅞	4⅛
13. Height of rear seat above floor	15¼	10¾	4½	17¾	8½	9¼
14. Breadth of rear-seat back:						
Full	64	50¼	13¾	54½	32¾	21¾
Split	19	19	0			
15. Height of rear-seat back	24½	19⅜	5⅛	24½	16½	8

Seat Position in Workspace

16. Horizontal adjustability	5½	3	2½	8	3	5
17. Rear-seat pan to roof	37¼	25¾	11½	36	18½	17½
18. Rear-seat back to front-seat back	44½	22	22½	33¼	17	16¼
19. Top of rear-seat back to rear shelf	3	2½	5½	15	½	14½
20. Front-seat pan to roof	37¾	34	3¾	40	34½	5½
21. Front-seat back to instrument panel	38⅜	26	12⅜	38½	27½	11
22. Seat pan to lower edge of steering wheel	6¾	4½	2¼	8	5	3
23. Seat back to lower edge of steering wheel (forward)	14¼	9¾	4½	14½	9½	5
24. Seat back to lower edge of steering wheel (rear)	19	14¼	4¾	21½	14	7½
25. Top front of seat to clutch pedal	19	18½	½	24	17	7
26. Top front of seat to brake pedal	22	18⅛	3⅞	24	16⅝	7⅜
27. Top front of seat to accelerator pedal	23¼	20¾	2½	25	18½	6½
28. Top front of seat to dimmer switch	29	22	7	28¼	19½	8¾

Steering-wheel Position in Workspace

29. Steering wheel to left door	6¾	2⅞	3⅞	6	2	4
30. Steering wheel to brake pedal	27¼	23½	3¾	26	23	3
31. Steering wheel to accelerator	30	27	3	29	25⅛	3⅞
32. Steering wheel to clutch pedal	25¼	23½	1¾	26¼	22½	3¾
33. Steering wheel to dimmer switch	33¼	30½	2¾	30¾	26¾	4
34. Steering wheel to emergency-brake pedal (on)	31⅞	27½	4⅜			
35. Steering wheel to emergency-brake pedal (off)	27¾	24½	3¼			
36. Steering wheel to floor	18⅜	13	5⅜	20⅜	13½	6⅞
37. Acute angle of steering wheel	80°	56°	24°	79°	41°	38°

Steering Wheel

38. Steering-wheel diameter	18	16½	1½	17⅜	15½	1⅞
39. Depth from outer rim to steering column	4	0	4	3⅜	−1	4⅜
40. Diameter of steering-wheel rim	⅞	¾	⅛	⅞	⅝	¼
41. Number of spokes	4	2	2	4	1	3
42. Horn-ring diameter	12¾	3	9¾	12	1⅜	10⅝

TABLE 14-3 *Comparison of Selected Dimensions of the Driver's Workspace in American and Imported Automobiles (1959 Models) (Continued)*

	Domestic vehicles (N = 20)*			Imported vehicles (N = 22)*		
	Maximum	Minimum	Range	Maximum	Minimum	Range
Foot Controls						
43. Width of accelerator	2	1 3/8	5/8	2 1/4	1	1 1/4
44. Angle of accelerator	124°	105°	19°	155°	100°	55°
45. Doghouse to accelerator	3 1/4	1 1/2	1 3/4	3 1/2	3/4	2 3/4
46. Brake pedal to accelerator	4 3/4	1 5/8	3 1/8	4	1 1/4	2 3/4
47. Height of brake pedal	7 3/4	4 3/4	3	8	4 1/4	3 3/4
48. Brake to clutch pedal	3 1/2	2	1 1/2	3 1/2	1	2 1/2
49. Brake pedal to steering column	7 1/2	1 5/8	5 7/8	8 7/8	5/8	8 1/4
50. Width of brake pedal	8	3	5	3	1 1/2	1 1/2
51. Angle of brake pedal	150°	104°	46°	140°	84°	56°
52. Height of clutch pedal	8 1/8	5 5/8	2 1/2	8	4 1/4	3 3/4
53. Side wall to clutch pedal	8 1/2	4 7/8	3 5/8	9 3/4	3/4	9
54. Width of clutch pedal	3 1/2	3	1/2	3	1 1/2	1 1/2
55. Angle of clutch pedal	115°	104°	11°	140°	84°	56°
56. Clutch pedal to steering column	6 3/4	3 5/8	3 1/8	10	1 1/2	8 1/2
57. Dimmer switch from left wall	3 1/2	3/4	2 3/4	4 3/8	1	3 3/8
58. Dimmer switch from clutch	8	3	5	9	0	9
Anthropometric						
59. Steering wheel to vent-window crank	11 1/2	4 1/4	7 1/4	9	2	7
60. Steering wheel to window crank	18 1/2	6 1/8	12 3/8	19 1/2	6 1/4	13 1/4
61. Steering wheel to door crank	14	3 5/8	10 3/8	10 1/2	2 3/4	7 3/4
62. Steering wheel to ignition switch	14 1/4	7 3/4	6 1/2	16	6	10
63. Steering wheel to starter switch	14 1/4	7 1/4	7	13 3/4	4 3/4	9
64. Steering wheel to headlight switch	13	6	7	12 3/4	3 1/8	9 5/8
65. Steering wheel to instrument-light switch	13	6	7	14 3/4	3 1/8	11 5/8
66. Steering wheel to seat adjustment	28	18 3/4	9 1/4	27	16	11
67. Steering wheel to temperature control	21 3/4	8	13 3/4	23 3/4	7 1/4	16 1/2

68. Steering wheel to cigarette lighter	20¾	8	12¾	17¼	7¾	9½
69. Steering wheel to ash tray	21	7¾	13¼	16¾	8	8¾
70. Steering wheel to windshield wiper	13⅜	6¾	6⅝	15⅝	5½	10⅛
71. Steering wheel to choke	12¼	11¼	1	18¾	6½	12¼
72. Steering wheel to glove-compartment button	31¾	14	17¾	24	12	12
73. Steering wheel to emergency brake (off)	19¾	14	5¾	24½	15	9½
74. Steering wheel to emergency brake (on)	18¼	11	7¼	19¼	11	8¼
75. Steering wheel to map-lights switch	41¼	6	35¼	46⅜	12¾	33⅝
76. Steering wheel to blower switch	21½	7⅝	13⅞	17¾	4¼	13½
77. Steering wheel to defroster switch	21½	7⅞	13⅝	24½	5½	19
78. Steering wheel to clock	39½	7¾	31¾	16⅛	8½	7⅝
79. Steering wheel to air-vent control	22	7	15	32	5	27
80. Steering wheel to radio control (on-off)	19	11½	7½	21¾	6	15¾
81. Steering wheel to heater switch	20⅜	7½	12⅞	20½	4½	16
82. Steering wheel to left sun visor	21	16	5	20	16	4
83. Steering wheel to rear-view mirror	19	8	11	17½	7½	10
84. Steering wheel to horn	7½	2	5½	8	2⅜	5⅝
85. Steering wheel to directional-signals control	9	2½	6½	10½	2½	8

* All data are in inches unless otherwise specified.

VARIATION IN GEOMETRY OF WORKSPACE AND PASSENGER SPACE OF 20 DOMESTIC AND 22 FOREIGN PASSENGER CARS

In the analysis of vehicles manufactured during 1959, 84 specific measurements were obtained. These were classified into seven general categories:

1. Windshield areas
2. Seat dimensions
3. Seat position in workspace
4. Steering-wheel position in workspace
5. Steering-wheel dimensions
6. Foot-control dimensions and position in workspace
7. Arm and leg reaches

Standard procedures were used for the collection of vehicle measurements.

The comparative ranges of measures between and among domestic and foreign vehicles are given in Table 14-3. Because the data included in this table are so extensive, only those variables for which sizing criteria have been developed will be discussed.

Windshield Area (see Table 14-3, items 1 to 4)

By itself windshield area is not necessarily a significant variable, since visibility depends also upon (1) eye height, (2) distance of the eye from the windshield, (3) obstruction to vision, and (4) the light-transmission qualities of the windshield. However, there is no significant correlation of these variables with the distribution of windshield areas in either foreign or domestic vehicles. Thus, windshield size and area of windshield cleared appear not to have been determined by any reference to biotechnical data except the necessity of providing some relatively transparent medium for vehicle windows.

For instance, both windshield area and range among foreign vehicles are less than among domestic vehicles. In foreign vehicles the maximum area was found to be 1,308 in.^{2} and the minimum 444 in.^{2}, about one-third the maximum area. In domestic vehicles the minimum area was found to be 894 in.^{2}, about one-half the maximum area of 1,888 in.^{2} Windshield wipers varied in length and range of function and thus cleaned different amounts of windshield area. Areas cleaned varied widely and were generally a small portion of the total area of the windshield. The largest wiper did not necessarily clean the greatest area. Thus there was measurable unconformity among area of windshield, length of wiper blade, and area of windshield cleaned. Even these relatively crude data allow positing the generalization that design specifications

could not have been generated by scientific data regarding the vision requirements of driver populations.

Seat Dimensions (see Table 14-3, items 5 to 9, 11 to 15)

Breadth of Seat. Seat breadth is generally measured by buttock breadth. However, when several persons sit side by side, elbow breadth, a larger measurement, has been found to be a better datum. For an individual seat the recommended minimum breadth is 18 in. All vehicles examined, both imported and domestic, met or exceeded this minimum for individual seats. For seating three persons side by side, a minimum seat breadth of 58 in. has been recommended. Of domestic vehicles, 11 failed to meet the suggested minimum requirement, and, of the imported cars, none met the minimum for three persons. The range of seat breadths of these imported vehicles was 44½ to 50¼ in., a range of 5¾ in., with all measurements well under the suggested minimum. The domestic vehicles varied from 55¼ to 60 in., a range of 4¾ in. Rear seats varied more than front seats. In domestic vehicles the rear-seat breadths ranged from 52 to 66¼ in. In four of the vehicles examined, rear-seat breadth was less than the suggested minimum. The rear-seat breadths of imported vehicles ranged from 34¼ to 57½ in. Again, no imported vehicle met the suggested minimum seat breadth for three persons seated side by side.

Depth of Seat. The recommendation of 18 in. for depth of seat was based on the buttock-to-popliteal measurement.[2] In domestic vehicles the seat depth varied from 19¾ to 17 in. for the front seat and 20 to 16 in. for the rear. Only two domestic vehicles met the requirement in the front seat, but seven more were between 18 and 19 in. The variation was greater in the rear seat, where the dimension is not so critical. Imported vehicles ranged from 21 to 17 in. in the front and 20 to 15⅞ in. in the rear. Only three vehicles were in the 18- to 19-in. range, and only one met the requirement. In this dimension, there should be little variance. If the seat is too shallow, there is insufficient support; if too deep, the buttock must be shifted forward to an uncomfortable position in order to prevent chafing the rear of the calf.

Height of Seat Back. The recommended height is no less than 18 or more than 21 in. Of the domestic vehicles examined 1 fell below the recommended minimum, and 4 were over the recommended maximum in the front seat. None fell below the minimum in the rear seat, and 10 were over the maximum. However, since the maximum was suggested to prevent providing a headrest for the driver, this dimension is not at

[2] The distance from rear of buttock to rear of lower leg immediately under knee joint at 90° angle.

all critical in the rear seat. In the imported vehicles, 1 was under the minimum and 3 over the maximum recommended for the front seat, and there were 2 under and 10 over for the rear seat. Once again the variation was high: 4¾ in. for the front-seat back of domestic vehicles and 5¼ in. for the front-seat back of imported vehicles. Again the rear seat varied more than the front seat: 5⅛ in. range in the domestic vehicles and 8 in. in the imported.

Breadth of Seat Back. Fifty-eight inches is the minimum seat-back breadth recommended for accommodating three persons. Fifty-four inches is suggested as an "acceptable" minimum. No front-seat back of the three-passenger seat was as wide as the suggested minimum. However, in the domestic vehicles, all those with individual seats were above the recommended 20-in. minimum. Of the 18 imported vehicles with individual seats, 7 had seat backs less broad than the recommended 20-in. minimum. Once more, the rear-seat backs varied more than the front. The range was 13¾ in. in domestic vehicles and 21¾ in. in imported vehicles, from 50¼ to 64 in. and from 32¾ to 54½ in. for the breadth of the rear-seat back. The recommendation is based on buttock breadth, a smaller dimension than elbow breadth. More study is necessary before definitive statements can be made for these measurements. However, the variation found among the vehicles examined is considerably more than should be expected.

Seat Height. This measurement is based on popliteal height, and the recommendation is an optimal range of 10 to 14 in. Of the domestic vehicles examined, eight were under and none was over the minimum suggested for the front seat. Rear seats were generally higher than front seats. None of the domestic vehicles examined had rear seats lower than the suggested minimum, but only two exceeded it. Four of the imported cars examined were under the minimum optimal recommended height for the front seat, and one was under for the rear seat. Two were over the 14-in. maximum for the front, and seven were over for the rear. The range in the imported vehicles was 9¼ in. (8½ to 17¾) for the rear and 6⅝ for the front seat. The domestic vehicles had less variation (5¼ in. front and 4½ in. rear) but still showed a rather wide range. It is somewhat pertinent to point out that nearly half the American female-driver population does not have an eye sitting height greater than the height of the upper edge of the steering wheel.

Steering-wheel–Seat Dimensions (see Table 14-3, items 22 to 24)

Seat–Steering-wheel Distance. The minimum recommended distance is 7 in. More clearance should be provided if possible, but not at the expense of raising the steering wheel to an uncomfortably high position. No domestic vehicle examined and only three of the imported vehicles

provided the recommended minimum clearance. The measurement was made to the undeflected cushion at the front of the seat, where deflection is minimum. Domestic vehicles ranged from 4½ to 6¾ in., all below the recommended minimum. The deflection of the cushions ranged from 1 to 2 in. *at point of maximum deflection,* directly beneath the buttocks, which is not the same point as that at which the distance between steering wheel and seat pan was measured. Thus, full deflection cannot be included in this measurement. If 1 in. is allowed for deflection (this varies with seat position, since, the farther back the seat, the less the deflection that should be allowed), then seven domestic vehicles may be said to have met the minimum recommendation. This, however, varies greatly with operator somatotype and seat position, as has been stated. Better standards are necessary before definitive statements can be made.

Back-rest–Steering-wheel Distance. The relevant measurement is abdomen depth. This skewed body measurement is one that is extremely widely distributed in the population and varies independently from other body dimensions. The recommended minimum distance is 14 in. when the seat is in the central position of its fore-and-aft adjustment range. Thirteen of the domestic vehicles and thirteen of the imported vehicles examined had less than the recommended minimum when they were in the center of the adjustment range. In the imported vehicles the variance was from 12¼ to 16½ in., a range of 4¼ in. However, with the seat fully forward, only one domestic vehicle and two imported vehicles had the recommended 14-in. clearance.

Seat-cab Dimensions (see Table 14-3, items 17, 18, 20, and 21)

Back-rest–Dashboard Dimension (in rear seat; back-rest–back-of-front-seat dimension). This measurement is taken from the back rest to the nearest forward obstruction at knee level in the vertical fore-and-aft plane of the knee. The recommended minimum is 27½ in. Only 1 domestic vehicle was less than the recommended minimum in the front seat. No imported vehicles examined were below this minimum. Nine domestic vehicles were under the suggested minimum for the rear seat, and only 2 of the 20 foreign vehicles with rear seats met the minimum requirement. The measurement was always a minimum, for the front-seat measurement was taken with the front seat forward and the rear-seat measurement was taken with the front seat in its rearmost position. Once again the range was wider for the rear seat. In this dimension the space in the rear seat was generally somewhat smaller than that in the front seat, whereas in other dimensions (seat breadth, height, depth) it was usually larger. In the imported vehicles this dimension was sometimes so small that entering and leaving the vehicle was made fairly difficult.

Front-seat Pan to Roof. In all cases but one, this measurement was taken with the seat deflected by model 1, 182 lb, 68th percentile (see Table 14-4). The exception was one measurement with model 2, 159 lb, 35th percentile. The recommended minimum height is 40 in. This recommendation is meant as a minimum; more space is desirable. Eleven of the domestic vehicles and ten of the imported vehicles did not meet this minimum requirement in the front seat. Once again the rear seat, in which the dimension is not so critical, varied in a far greater range than the front seat—from much less to slightly more space than the minimum requirement in the domestic vehicles and from very much less to not quite as much head space in the imports. The seats were also measured without accounting for cushion deflection. Without the deflection caused by the models, no imported vehicles met the suggested minimum. This means that (1) the seat-roof dimension is less when the operator is entering and leaving the car and (2) people who weigh less than model 1 will have less head space. A person actually needs more space when in motion than when at rest. This suggests that space is lessened when it is most needed, when the passenger is entering or leaving the vehicle.

Distance from Steering Wheel (or Shift Lever) to Brake Pedal (see Table 14-3, items 30 and 31). This dimension is based on the actual vertical distance from the floor to the top of the knee, with both knee and ankle forming right angles. The recommended minimum distance is 24 in. This again is a minimum; more space should be provided, but not at the expense of raising the steering wheel to an uncomfortable height. The domestic vehicles examined generally met this minimum. One vehicle fell below the minimum. The variation was from 23½ to 27¼ in., a range of 3¾ in. Eight imported vehicles were below the suggested minimum. The variation in these was from 23 to 26 in., a range of 3 in. All these figures are for the brake pedal. Only three of the domestic vehicles had clutch pedals. In one of these the distance was less than the suggested minimum, making that statistic almost valueless. Of the imports, all had clutch pedals, and the distance to the pedal was generally the same as that to the brake, making the clutch dimension unnecessary.

Seat Adjustability (see Table 14-3, item 16)

Vertical Adjustability of the Front-seat Structure. Almost no data were obtained, since vertical seat adjustability was lacking except for those vehicles equipped with power seats. No rear seat was found that was adjustable either horizontally or vertically. No imported vehicle had vertical seat adjustment. Only domestic vehicles with powered seats had

vertical adjustment, although members of the research team were told that the standard seat could be mounted on blocks to raise it permanently. Of those seats with adjustment, the range was ¼ in., from 1¾ to 2 in., without increments (i.e., in a continuum from lowest to highest point). The recommended vertical adjustment is 4½ in. in increments of ½ in. or less.

Horizontal (Fore-and-Aft) Seat Adjustability. The recommended minimum adjustment is 8 in. in increments of ½ in. or less. The reason for this seemingly large range of adjustability is that it is "functional leg length" that is the relevant body measurement. This is the distance from the most posterior point of the buttocks to the surface of the outstretched foot while the body is in seated position. The knee angle varies widely, and thus the distance varies widely. The recommended range is based upon a composite measurement of seat height vs. popliteal height, seat depth vs. buttock-popliteal length, thickness of thigh, seat deflection, and other variables. The range of horizontal adjustability in domestic vehicles examined was 2½ in., from 3 to 5½ in. No domestic vehicle came near the suggested minimum. One imported vehicle had 8 in. of adjustability, and one had 7½ in. The rest ranged from 3 to 5 in. All vehicles examined had adjustability over their range in increments of ½ in. or less.

Angle of Steering Wheel (see Table 14-3, item 37)

Kephart and Dunlap [4] in 1955 experimentally derived a range for steering-wheel angles of 42.51 to 48.17°, with a preferred angle of 45.34°. This range was developed on the basis of seat heights varying between 14.69 and 17.39 in., which had been experimentally derived from subjects' selecting seat heights and wheel angles individually. The measured steering-wheel angle of the domestic vehicles examined ranged from 56° to 80°25′, a range of 24°25′. The wide variance in both steering-wheel angle and seat height was unexpected, since there exist empirical scientific data for guiding design. No domestic vehicle examined was within the criterion derived by Kephart and Dunlap. Imported vehicles varied more than the domestic vehicles, from 78°50′ to 41°20′, a range of 37°30′. All vehicles except one exceeded the Kephart-Dunlap standard.

The greater angles tend to favor the driver who must place his seat toward the rearmost limit of adjustment, for, the farther back the operator sits, the greater the probability that he will prefer the steering wheel that approaches the vertical aspect. It can be seen that many of the other categories of workspace dimensions were more appropriate for the driver of smaller stature, whereas this distribution of steering-wheel angle measures was more appropriate for drivers of larger stature.

The actual relations of seated drivers to the locations of seat and steering wheel may be seen in Figure 14-1*a* to *c* for stock domestic automobiles. The physical dimensions of the two human models shown are presented in Table 14-4.

TABLE 14-4 *Anthropometric Measurements of Models**

Anthropometric variables	Model 1		Model 2	
	Measurement	Percentile	Measurement	Percentile
1. Weight, lb	182	68	159	35
2. Height	72.5	94+	67.5	36
3. Abdominal depth	9.6	28	8.7	10
4. Arm span, total	78.1	....	68.4	
5 Buttock-knee length	25.0	89	22.8	63
6. Calf circumference	14.7	....	14.4	
7. Chest breadth	12.0	....	11.9	
8. Chest circumference	39.8	....	37.2	
9. Chest depth	9.5	....	9.1	
10. Elbow breadth	17.6	49	17.0	35
11. Elbow span	40.4	....	36.8	
12. Elbow–middle-finger length	21.0	....	18.0	
13. Foot breadth	4.1	80	3.8	29
14. Foot length	11.4	97	9.9	16
15. Hand breadth	3.7	....	3.2	
16. Hand length	8.4	....	7.0	
17. Head circumference	23.0	....	22.0	
18. Hip breadth	14.3	....	13.8	
19. Knee breadth	8.2	....	8.0	
20. Knee height (lower-leg length)	23.4	94	20.5	16
21. Shoulder breadth	18.8	86	18.2	66
22. Shoulder-elbow height	17.0	....	13.8	
23. Sitting height, erect	36.8	71	35.5	36
24. Sitting height, normal	35.7	....	33.6	
25. Trunk height	24.2	82	22.6	31

* All data in inches unless otherwise specified.

The figure shows models 1 and 2, in order of size, seated in the workspaces of the vehicles, with the seat of each vehicle in the lowest and rearmost position. It can be seen that the larger of the models, No. 1, has barely sufficient space to raise his foot from the accelerator to the foot brake, regardless of the type and make of car. In every instance his thigh is depressed by the lower rim of the steering wheel. These models are not exceptionally outsized; the lower-leg length of the larger model is in the 94th percentile, and that of the smaller of the models is in the 16th percentile. Shorter operators may take advantage of the horizontal adjustment of the seat, which is useless for the larger

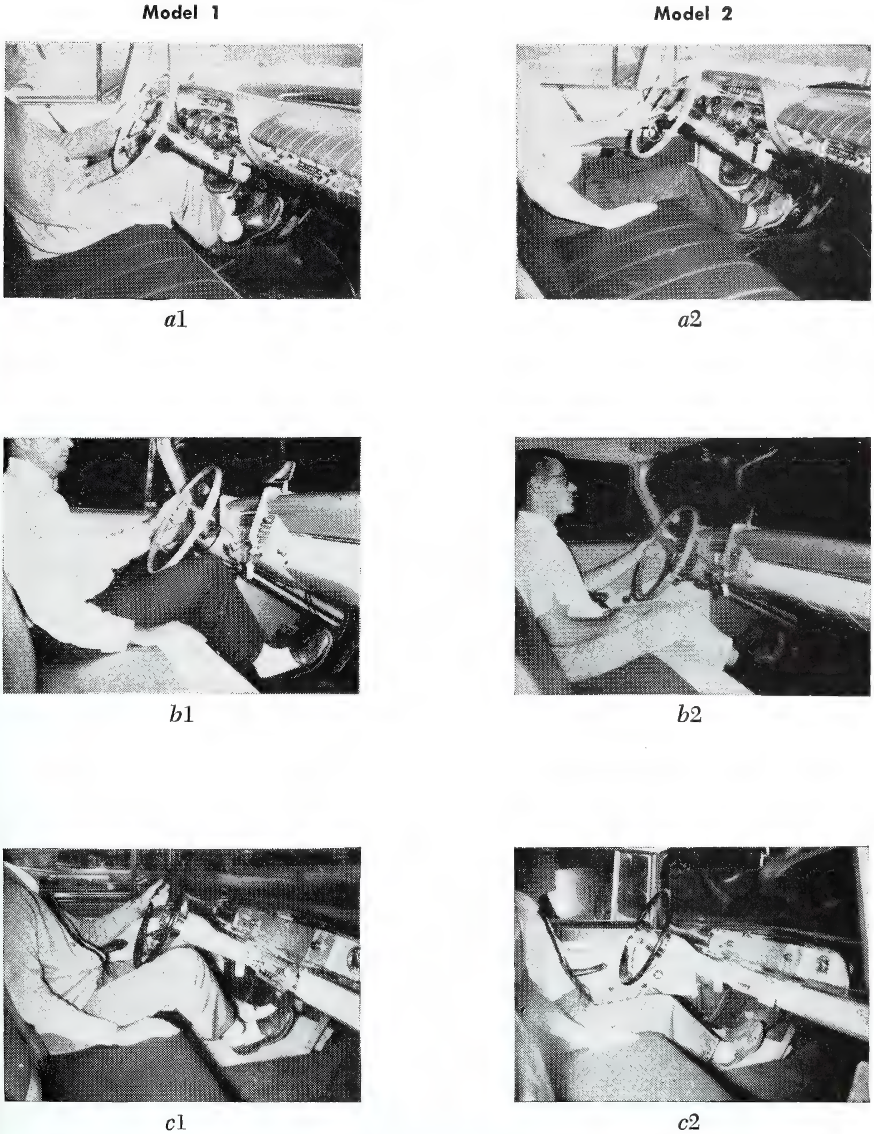


FIG. 14-1 *Relation of seated drivers to the steering wheel in samples of domestic and imported automobiles.*

driver. For the taller operators, the entire seat in some vehicles may be detached and the seat frame removed and repositioned. This procedure, however, results in reducing rear passenger space.

As has been demonstrated, there are measurable differences between the systematic data developed by Stoudt, Damon, and McFarland [16]

and those developed by the automobile industry. There is also disagreement over how much of the driver population should be taken into consideration when the design geometry of space is determined. For example, at a recent conference, the Society of Automotive Engineers approved designing the workspace to accommodate the lower 90 per cent of the population [15]. This position indicates that no special design provisions should be made for the larger driver. This is an untenable design criterion, for it is not necessarily true that larger operators can be accommodated in vehicles designed for smaller operators. In fact, it is clear that using the lower 90 per cent of the distribution would accommodate the extremely small person and disaccommodate the extremely large one, thus favoring one end of the distribution at the expense of the other.

Machine workspace must accommodate both ends of the distribution. Designing for one end does not automatically lead to adequate arrangements for the other. The consequences of arbitrary design criteria are not difficult to assess. For example, Table 14-1 clearly shows that almost all female anthropometric distributions fell in the lower half of the equivalent male distributions. Imposing the 90th percentile criterion means penalizing males who are significantly larger than females. In 1958 there were 30.5 million female license holders and 50 million male license holders. Therefore, design penalty would be imposed upon a total of 10 million male drivers and an unknown number of female drivers. It would be a simple matter to calculate the consequence of any other arbitrary cutoff points.

Since there is no agreement as to which criterion or criteria may be selected, it is suggested that imposition of any particular design requirement is fallacious, for there is no evidence that a design cannot be developed that would accommodate all the driver population, perhaps with the exception of anthropometric oddities. The motivation for other than full accommodation of the driver population should be examined carefully.

BIBLIOGRAPHY[3]

1. Damon, A., and R. A. McFarland: The Physique of Bus and Truck Drivers: With a Review of Occupational Anthropology, *Am. J. Phys. Anthropol.*, vol. 13, pp. 711–742, 1955.

2. Haber, H.: The Safety Hazard of Tinted Automobile Windshield at Night, *J. Opt. Soc. Am.*, vol. 45, no. 6, pp. 413–419, 1955.

[3] With the exception of items 2 to 5 and 15, all the projects referenced here were sponsored by the Commission on Accidental Trauma of the Armed Forces Epidemiological Board, Department of Defense.

3. Heath, W. M., and D. M. Finch: Effects of Tinted Windshield Glass on Night Time Visibility Distances, *Highway Research Board Bull.*, no. 68, pp. 1–15, Jan. 13, 1953.

4. Kephart, N. C., and J. W. Dunlap, Jr.: "Human Factors in the Design of Vehicle Cab Areas," Occupational Research Center, Purdue University, Lafayette, Ind., 1955.

5. Lauer, A. R., et al.: Effects of Night Glasses and Colored Windshields, *Optometric Weekly*, vol. 41, no. 25, pp. 951–955, 1950.

6. McFarland, R. A., A. Damon, and H. W. Stoudt: Anthropometry in the Design of Driver's Workspace, *Am. J. Phys. Anthropol.*, vol. 16, no. 1, pp. 1–24, 1958.

7. ———, ———, and ———: The Application of Human Body Size Data to Vehicular Design, *SAE Spec. Publ.* 142, 1955.

8. ——— and R. G. Domey: Biotechnical Aspects of Driver Safety and Comfort, *Trans. SAE*, vol. 66, pp. 630–648, 1958.

9. ——— and ———: Human Factors in the Design of Passenger Cars: An Evaluation Study of Models Produced in 1957, *Proc. Highway Research Board*, vol. 39, pp. 565–582, 1960.

10. ——— and H. W. Stoudt: Human Body Size and Passenger Vehicle Design, *SAE Spec. Publ.* 142A, January, 1961.

11. ——— et al.: Dark Adaptation as a Function of Age and Tinted Windshield Glass, *Highway Research Board Bull.*, no. 255, pp. 47–56, January, 1960.

12. ——— et al.: "Human Body Size and Capabilities in the Design and Operation of Vehicular Equipment," Harvard School of Public Health, Boston, 1953.

13. ——— et al.: Human Factors in the Design of Highway Transport Equipment: A Summary Report of Vehicle Evaluation, Harvard School of Public Health, Boston, 1953.

14. ——— et al.: Human Factors in the Design of Trucks: An Evaluation Study of Models Produced in 1956, Harvard School of Public Health, Boston, 1957.

15. Preprint, A2-D Manikin—The Inside Story, *SAE Spec. Publ.* 267A, 1961.

16. Stoudt, H. W., A. Damon, and R. A. McFarland: Heights and Weights of White Americans, *Human Biol.*, vol. 32, no. 4, pp. 331–341, December, 1960.

17. Wolf, E., R. A. McFarland, and M. Zigler: Influence of Tinted Windshield Glass on Five Visual Functions, *Highway Research Board Bull.*, no. 255, pp. 30–46, January, 1960.

Chapter 15

CRASH DECELERATION TESTS WITH HUMAN SUBJECTS

*James J. Ryan**

AT THE PRESENT TIME, AUTOMOBILES ARE CONSTRUCTED ON THE PREMISE that "they are not built to crash"; yet in 1960 there were over 10,000,000 accidents [1]. In Minnesota alone, with a population of about 3,400,000 (1960 census), there were 723 deaths in traffic accidents, or about 2 each day [4]. This figure is even more startling when compared with the 337 deaths in commercial aviation during all of 1960 [2].

Upon collision, people riding in cars continue forward in a straight line, without change in position, until they strike a surface that pushes back hard enough to stop them. Automobile surfaces are not designed to arrest this motion safely. If all the deaths caused by people's striking the interior of the automotive structure, being thrown out through open doors, being impaled on a rearward-driven steering post, being jammed between levers and walls, or striking a sharp button, knob, spoke, or rim [12] were circumvented by safe design, the number of fatalities would be reduced by about one-half.

At the same impact speed, heavy cars require more force to stop them than light vehicles, since more energy must be dissipated. Although the energy increases with the square of the velocity, the front-end deformation is approximately proportional to the speed [10, 11]. Thus, the more the deformation, the greater the decelerating force. Hitting a bridge abutment or a rigid wall produces the greatest forces. Because of overlap of the rending parts, two cars colliding head on produce forces that are slightly less. Of course, if one of the cars is lighter in weight, the lighter receives the greater force. Since it is the forces offered by the resisting surfaces that injure and kill people, the best possible means for survival is the reduction of these forces to the minimum.

* Automotive Safety Research Project, Department of Mechanical Engineering, University of Minnesota, Minneapolis, Minn.

The work described in this chapter is the result of studies and tests performed under contracts with the U.S. Air Force and by a grant-in-aid from the U.S. Public Health Service.

Automotive design research has shown the feasibility of a reduction of the forces on the passengers with seat belts in automotive crash to approximately one-fourth of those presently suffered. This reduction is the result of the use of a self-tightening seat belt with proper elasticity, strength, and damping in series, the engineering development of a hydraulic shock-absorbing bumper, the protection obtained for the driver by controlling the deceleration of the rotating upper torso on the padded steering post, and the protection provided for the passenger by a free forward swing with clearance at the dash.

SEAT BELTS

Many parts of the human body are vulnerable to small forces, and therefore, even at low speeds, injuries and deaths may occur. One of the most effective preventives of the application of small forces is the seat belt. If supported by a seat belt, the pelvic bone structure is probably capable of resisting greater forces than any other portion of the human body. It has been amply demonstrated that the use of seat belts in automotive vehicles reduces crash injuries and deaths [3, 5, 10 to 12].

The use of seat belts by the general public has been resisted for several reasons. Seat belts are in the initial stages of development and require custom purchasing and fitting. Their cost varies widely according to the type and the manner of attachment to the vehicle. In addition, if seat belts are restrictive to the individual, if they are caught in the door, are hard to fasten, or get dirty and discolored easily, it is more likely that people will sit on them rather than use them. Furthermore, the collision-force resistance provided by present seat belts is four times as great as is necessary. During collision the action of a body on a seat belt is like that of a weight on a spring when acceleration is suddenly applied to the spring. The stopping force increases with the passenger's movement into the seat belt. As a result, the application of the decelerating force to the wearer of a seat belt determines the loading condition. If the force could be controlled by hydraulic shock-absorbing bumpers on the vehicle and proper dashpots in series with the seat belt, the forces on the human would be considerably reduced [10 to 12].

Other factors affecting the value of seat belts are the elasticity of the webbing; the slack of the belt at the floor fastening and around the waist of the wearer; the seat movement, both as a free-flying object with a resulting added load on the belt and as a support with a restricted displacement; the type, location, and strength of the floor fastening; the angle of the belt at the floor fastening; the previous use and freedom

from weak spots of the seat-belt system; the rotation of the upper torso and the lower-leg areas of the human body which strike the forward structure of the vehicle, including the steering wheel and shaft, the dash, the underdash pedals and levers, and the corner posts; the penetration of the driver's compartment by forward elements such as the engine, steering post, wheels, and hood; and the resulting elongation of the belt, which determines whether it adequately restricts the body within the limited space. The last factor is an extremely important one. Tests by the Automotive Safety Research Project have indicated that slackness in the seat belt increases proportionally the deceleration forces upon the wearer. That is, after impact the vehicle gradually slows down, whereas an unrestrained body continues to move ahead rapidly until it is stopped by the seat belt. Slackness or looseness of the seat belt is measured by the arbitrary adjustment of the belt about the rider's body, the flexibility of the mounting connection on the floor, and the straightening out of the belt to form the shortest line for supporting the body. The elongation of the belt usually does not increase in proportion to the force, but greater forces do cause greater elongations. Thus, with greater forces, larger clearances are required to prevent the body from striking interior surfaces.

To meet the many requirements of seat-belt effectiveness listed above, an automatic seat belt has been designed to be attached to the bar behind the seat. The holders on the bar have a light clock-spring mechanism in a reel that keeps the belt retracted to the rear of the seat and in proper position for immediate use. When the ends of the seat belt are grasped and slowly pulled forward, the belt may be easily fastened around the passenger with metal-to-metal clasps. If the belt is pulled rapidly, the reel moves forward up an inclination of 45° and teeth in the rim lock with matching teeth in the housing. The belt also locks by going up the incline when the vehicle is decelerated on impact. Cart tests in the Automotive Safety Research Project laboratory have shown an instantaneous "hold" on crash. The seat-belt holders may also be bolted to the floor in the same manner as the present seat-belt connections.

One advantage of such automatic seat belts is that, if the user rotates his body to back up the car or observe something at the rear of the vehicle or if he attempts to reach forward and touch the dash, the seat belt freely and lightly follows the body motions. The wearer is not restrained in forward, lateral, or rotary motion. If, however, a sudden movement is applied by the body on the seat belt, as in an accident, the belt is instantly locked and the passenger is restrained securely. When the buckle is released, the two ends of the seat belt retract to the back of the seat.

With this type of seat belt, means must be provided for anchoring

the seat securely in position. In present-day cars, seats are pulled loose from their tracks on impact by their own inertia and by the motion of the passengers, since the seats are not adequately supported and locked. Therefore, the seat is connected to the floor with flexible steel cables of sufficient strength to decelerate the passengers and yet allow any adjustments desired by the driver.

HYDRAULIC SHOCK-ABSORBING BUMPERS

Hydraulic shock-absorbing bumpers are designed to apply the maximum forces of retardation as early as possible in their displacement [9]. Thus the forces absorb the maximum energy at the highest velocities. The hydraulic shock-absorbing bumper, shown in Figures 15-1 and 15-2,

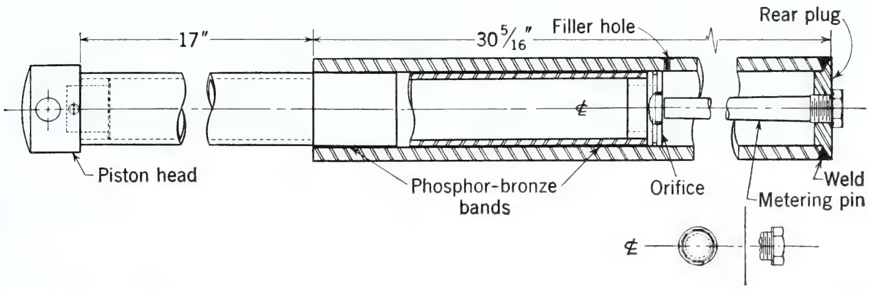


FIG. 15-1 *Outline drawing of hydraulic shock-absorbing bumper for a 4,000-lb passenger car—2¾-in. piston diameter, 1-in. orifice, 17-in. stroke.*

consists of a pair of steel tubes that telescope in such a manner that, when an axial force is applied, hydraulic fluid is forced from one tube to the other through an orifice whose area is controlled by the position of a metering pin. Thus, depending on the design of the metering pin, any rate of deceleration may be obtained with any size or weight of automobile or speed of impact. In terms of displacement and force, this hydraulic absorption of energy is the most efficient method known. With these bumpers the deceleration forces are lower for low-speed impacts and higher for high-speed impacts, whereas with metal-absorbing structures the deceleration forces are fixed.

In a 1957 crash demonstration, one of these bumpers was attached to the front of a 1956 Ford sedan, and, with the human driver and passenger held by seat belts, the car was driven into a rigid barricade at 20 mph. The deceleration forces on the hydraulic bumper, curve *A*, and the passenger compartment, curve *B*, are shown in Figure 15-3. The seat-belt forces during impact are shown as curve *C*. No injuries resulted to either the passengers or the car.

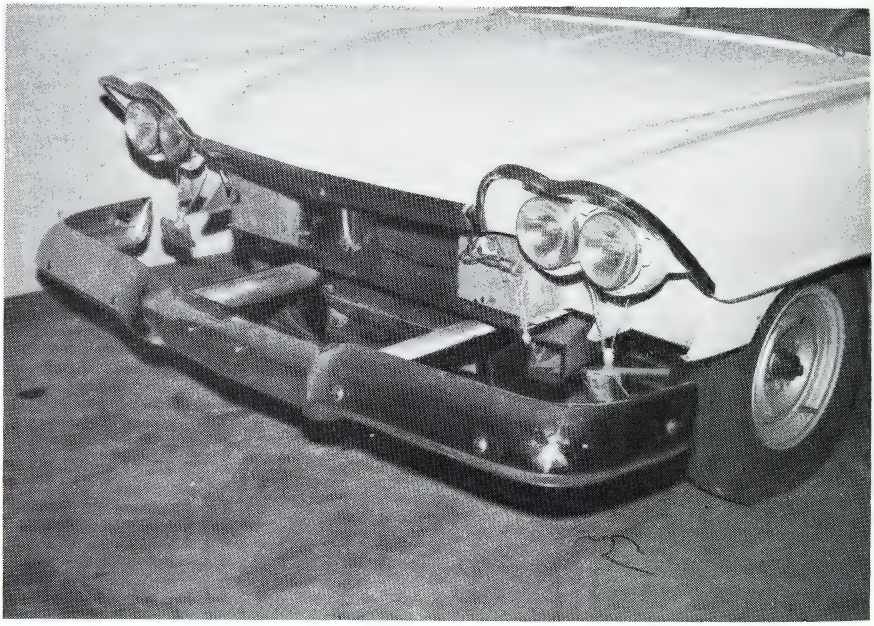


FIG. 15-2 *Hydraulic shock-absorbing bumper attached to front of 1959 Plymouth test car. (From Kenneth A. Roberts, "We Can Build a Crash-Proof Car!" Saga, vol. 23, no. 1, pp. 16–21, October, 1961. By permission of the publishers.)*

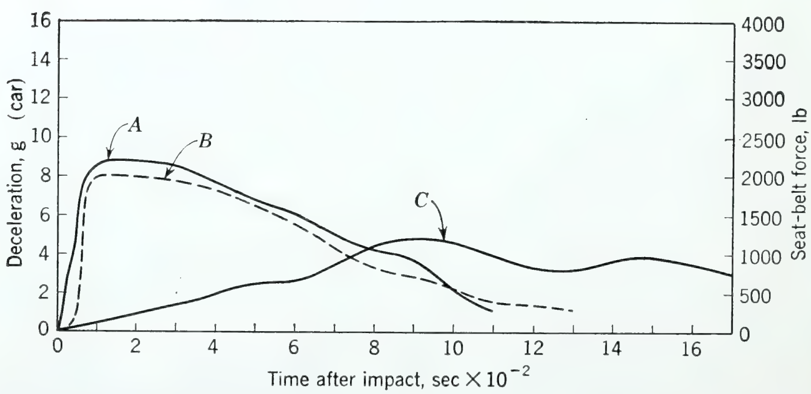


FIG. 15-3 *Deceleration forces for 1956 Ford with hydraulic bumper, reinforced frame, and recessed dash; 20-mph barrier impact with live passengers in seat belts. (A) Hydraulic-bumper deceleration (g); (B) passenger-compartment deceleration (g); (C) driver seat-belt forces (pounds).*

In Figure 15-4 are shown the relative impact decelerations as measured on a standard unmodified 1949 Chevrolet (curve *A*) and a 1959

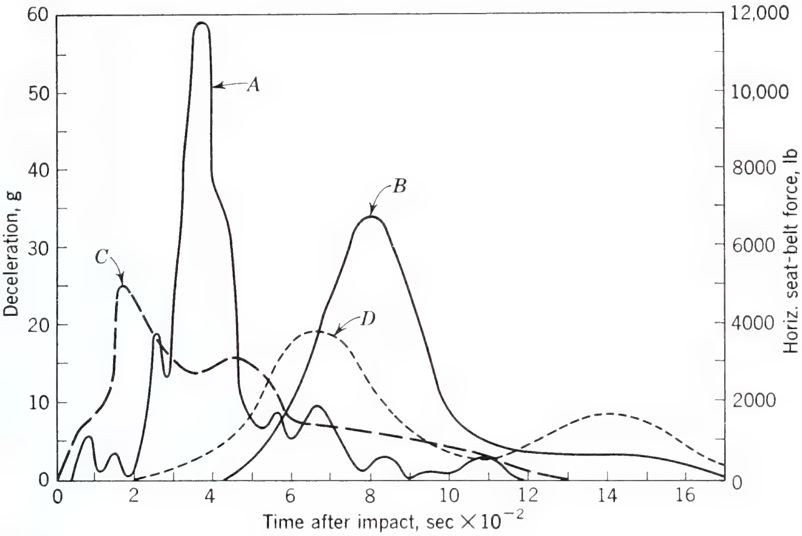


FIG. 15-4 *Impact decelerations of standard 1949 Chevrolet at 28 mph, (A) frame deceleration (g), (B) seat-belt forces (pounds), and 1959 Plymouth with hydraulic bumper at 30 mph, (C) frame deceleration (g), (D) seat-belt forces (pounds). (Curves A and B are drawn after D. M. Severy and J. H. Mathewson, Automobile Barrier Impacts, ser. 1, Institute of Transportation and Traffic Engineering, University of California, Los Angeles, Calif., 1956. By permission of the publishers.)*

Plymouth modified with a hydraulic shock-absorbing bumper (curve *C*). The respective seat-belt forces are shown as curves *B* and *D*. As can be seen, the hydraulic shock-absorbing bumper caused a significant decrease in frame deceleration as well as in the forces applied to the seat belt. Note also that there was an earlier build-up of force with the hydraulic bumper than with the standard bumper.

DECELERATION TESTS

A 2,000-lb steel cart was constructed that, when mounted on rails and propelled by a 10-in. air gun, obtained speeds of 15 to 40 mph. This cart was decelerated by an automotive hydraulic shock-absorbing bumper capable of applying a relatively uniform force over the stopping distance in order to absorb the maximum energy with the minimum force. An outline drawing of the cart is shown as Figure 15-5.

The first tests with the cart were made to determine the effect of

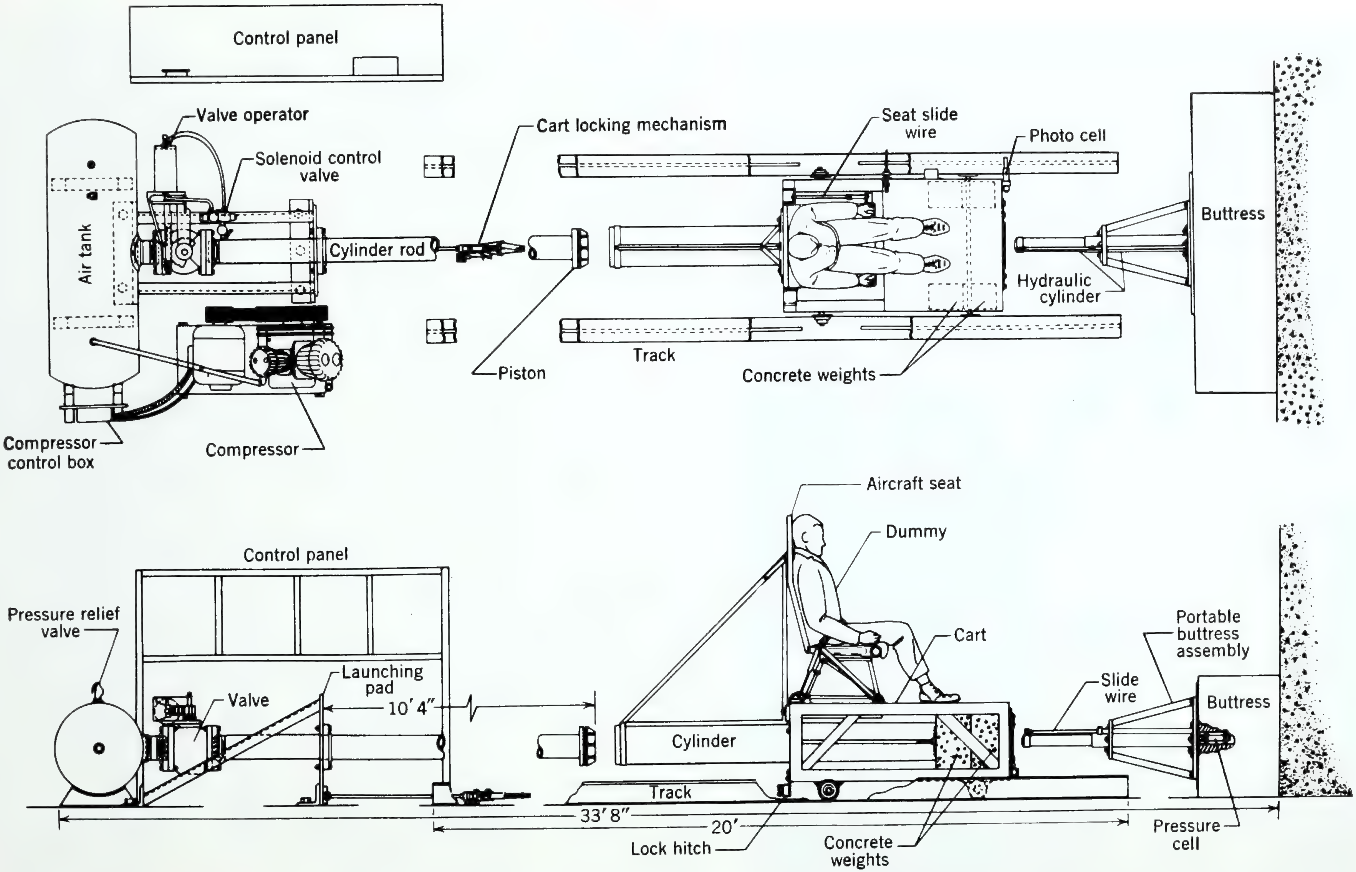


FIG. 15-5 *Assembly drawing of 2,000-lb cart with air-cylinder gun and hydraulic shock-absorbing bumper for bumper and seat-belt tests. (From J. J. Ryan and W. E. BeVier, Safety Devices for Ground Vehicles, Automotive Safety Research Project, University of Minnesota, Minneapolis, Sept. 1, 1960. By permission of the publishers.)*

changes in the design of the metering pin upon the characteristics of the hydraulic shock-absorbing bumpers. Oscillographic recordings of the bumper's internal pressure, cylinder stress, and piston displacement were made for a series of velocity impacts ranging from 15 to 40 mph and for cart loads of approximately 1,500, 2,000, and 2,500 lb. In addition, dummy riders were placed on the cart, and oscillographic and photographic measurements of deceleration, seat-belt pull, dummy displacements, and dummy decelerations were made. After a suitable hydraulic absorber was designed, seat-belt tests with dummies and human beings wearing multiple inelastic seat belts without slack were made in order to determine the seat-belt forces obtainable with proper engineering control. Results showing the horizontal seat-belt force corrected for angle of restraint from the rides of the author, weight 170 lb, at an impact speed of 16.6 mph, Peter A. Schoeck, weight 185 lb, at a speed of 18.5 mph, and the dummy, weight 200 lb, at a speed of 20 mph, were summa-

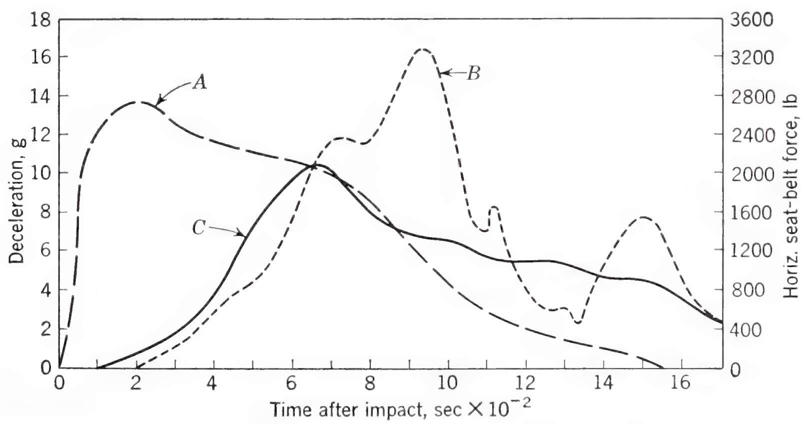


FIG. 15-6 *Decelerations and forces for Ryan, Schoeck, and dummy impacts corrected to impact velocity of 20 mph and 200-lb dummy weight. (A) Average cart deceleration (g); (B) average horizontal chest deceleration (g); (C) average horizontal seat-belt force (pounds).*

rized and averaged to an impact velocity of 20 mph and a 200-lb weight, as shown in Figure 15-6, curve *C*.

DYNAMIC SUPPORT IN A SEAT BELT

Determination of the equivalent weight for the dynamic support of a passenger in a seat belt requires consideration of the body dimensions. The pull of the seat belt is focused at the upper legs and the lower

part of the torso. The time of impact of a passenger in a seat belt is so short that during the impulse the body remains upright. Most tests have shown that at the instant of maximum seat-belt force the torso has not rotated past the vertical. Thus the horizontal seat-belt pull is approximately equal to the mass deceleration force on an equivalent portion of the body weight. The deceleration is measured on the upper legs (hip) in the forward direction.

The weight of a person on a seat belt is about 60 per cent of his total weight, since the rotation of the upper torso and lower legs requires less decelerating force. Thus, a deceleration of 20 g on a 200-lb man would result in a seat-belt pull of 20 times 60 per cent of 200 lb, or 2,400 lb.

The forces on a man in a seat belt respond in the same manner as in an idealized dynamic system. Therefore, consideration of the force input to the spring, the natural frequency of the system in terms of the input time, and the damping of the spring at the point of input may be demonstrated with respect to their variables. In Figure 15-7 are

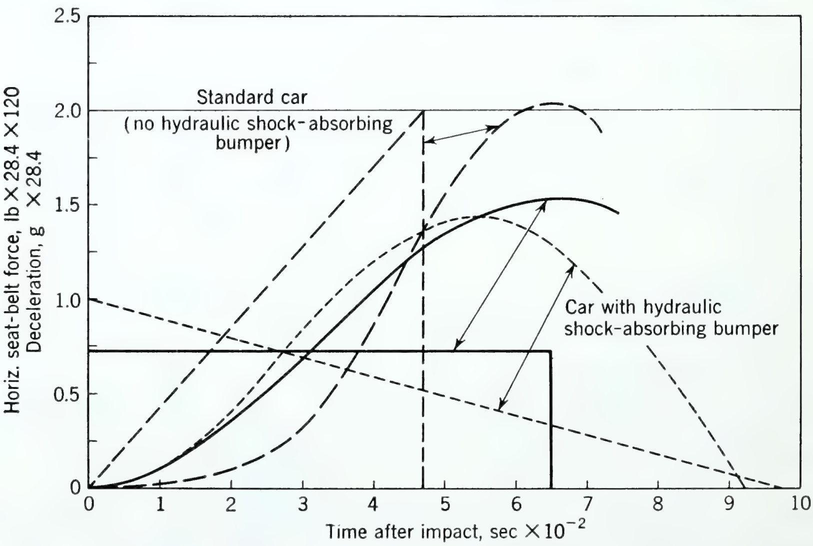


FIG. 15-7 *Hypothetical cart decelerations with rectangular and triangular bumper characteristics at 30 mph and corresponding horizontal seat-belt forces on a 200-lb man with an equivalent weight of 120 lb without damping. Deceleration distance 17 in.; seat-belt natural frequency 50 radians/sec.*

shown the hypothetical cart decelerations with rectangular and triangular bumper characteristics at 30 mph and the corresponding horizontal seat-belt forces on a 200-lb man with an equivalent weight of 120 lb without

damping. The decelerating distance is the standard 17 in., and the seat-belt system has a natural frequency of 50 radians/sec. For the same energy absorption, the triangular cart input with maximum force at the beginning results in smaller maximum seat-belt forces than with constant deceleration.

Figure 15-8 shows the same data, with the exception that the dynamic seat-belt system includes a damper in series that responds to 0.625 critical damping. It may be observed that in Figure 15-7 the standard maximum seat-belt force is slightly over 2, whereas in Figure 15-8, with

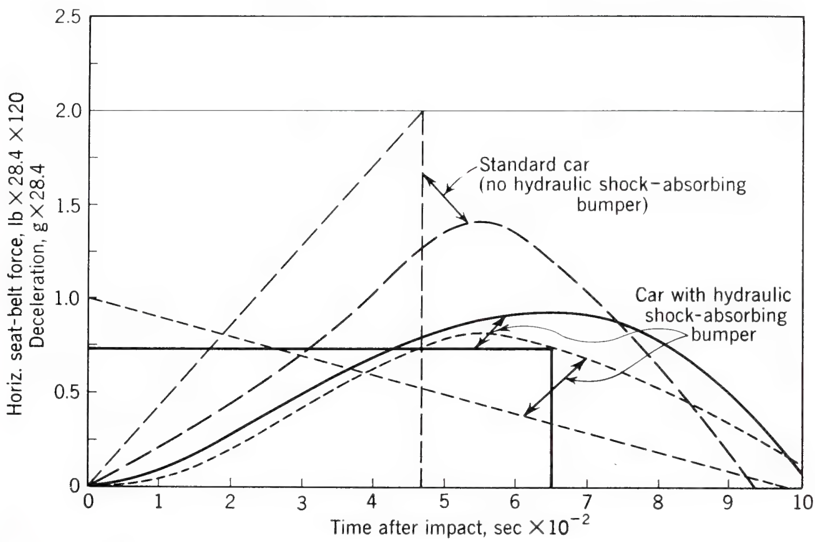


FIG. 15-8 *Hypothetical cart decelerations with rectangular and triangular bumper characteristics at 30 mph and corresponding horizontal seat-belt forces on a 200-lb man with equivalent weight of 120 lb with 0.625 critical damping. Deceleration distance 17 in.; seat-belt natural frequency 50 radians/sec.*

shock-absorbing bumpers and damping, the maximum seat-belt force is about 0.800. Thus the use of hydraulic shock-absorbing bumpers and seat belts with 0.625 critical damping would reduce the forces on the body to 40 per cent.

With a standard cart-deceleration input, the calculated decelerations acting upon a body with an elastic support and a dashpot damper in series are shown in Figure 15-9. Curves *A*, *B*, and *C* are for a natural frequency of 80 radians/sec and present values for damping ratios from infinity to 0.625 of critical. Curves *D*, *E*, and *F* are for the lower natural frequency of 50 radians/sec. It is observed that the extreme horizontal seat-belt forces are twice as great with a rigid dashpot and a natural

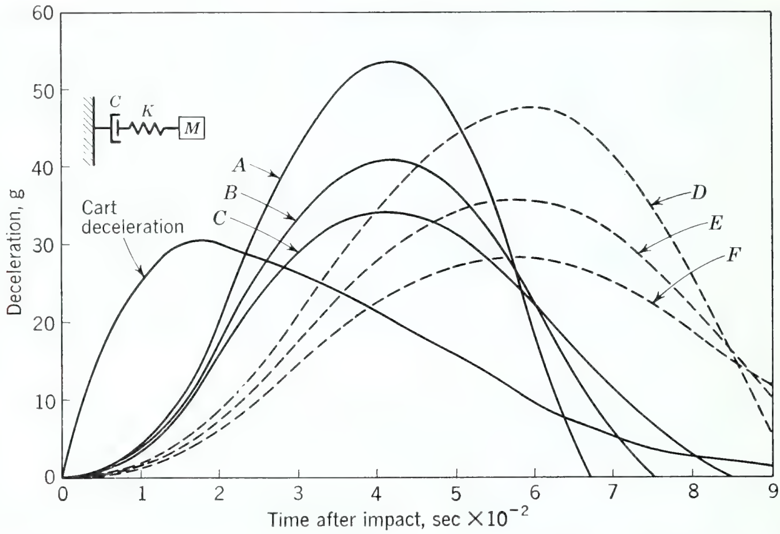


FIG. 15-9 *Horizontal pull of seat belt (in g) for equivalent weight of 120 lb on a 200-lb man.*

Curve	Natural frequency (radians/sec)	Critical damping ratio
A	80	Infinity
B	80	1.25
C	80	0.625
D	50	Infinity
E	50	1.25
F	50	0.625
	(Theoretical)	

frequency of 80 radians/sec (curve *A*) as for 0.625 critical damping and a 50-radians/sec elastic structure (curve *F*), the ratio being about 54:27 (G). These curves show the susceptibility of the man in the seat belt to respond in a fashion that may be applied to any similar dynamic system. Thus, the same mechanical factors that reduce the forces on a weight supported by a spring may be applied to the body in a seat belt.

PADDED STEERING WHEEL

After the free-swing seat-belt tests with Ryan and Schoeck, a number of runs were made with a steering wheel mounted in a normal position

in front of the dummy rider. It was found that, by automatic tightening of the seat belt and elimination of any movement of the seat, the chest of the driver could be positioned to strike the padded wheel on the steering post. Although the driver's head pitched forward and around the padding behind the wheel, his extremities were confined to a fairly restricted area. Tests indicated a slightly higher seat-belt load with the steering wheel in position but no secondary seat-belt pull at the end of the motion. The readings on the chest accelerometers were about the same for both steering-wheel and free-swing conditions.

A series of six runs was made in order to compare the effects of free forward swing and a padded steering wheel on seat-belt forces. It was indicated that a tightened seat belt contributed more to the reduction of the seat-belt forces than did the presence or absence of the steering post. However, the flailing extremities must have adequate clearance, or they must have built-in elements for providing the proper protection. Even when the seat belt was tightened (the maximum amount) before the test, the initial loading occurred between 2.5/100 and 5/100 sec after impact.

We can calculate the various displacements that are augmented by slack in the belt. If, for instance, the belt were slack for 2.5/100 sec, the cart would travel about 9 in. before the dummy tightened the belt; the slack would be about 4 in. Since the belt would have a greater load with a loss of deceleration distance, the elongation would be greater and would extend for a displacement of about 14 in. This belt force would be nearly twice that with no slack.

However, with a belt tightener connected to the hydraulic bumper system, not only did the seat belts begin to take load in less than 1/100 sec, but they also held an appreciable load after the system had come to rest. With the dashpot the pressure would be released during the loading period, and no residual force would exist. The final distance, the deceleration distance with tightened belt and dashpot, is perhaps the most correct of the assumptions discussed here: if the seat belt were tightened, the body would slow down with the cart, and there would be an additional distance of an assumed 3 in. for the dashpot to retract. These factors must be considered in positioning the individual in the seat belt and in determining the mechanical requirements for the minimum impact forces that will be carried by his body.

The seat-belt tightener is a piston and cylinder arrangement by which hydraulic pressure from the first 5 in. of the shock-absorbing-bumper stroke is partially bypassed in order to displace the piston and apply forces to the two ends of the seat belt. This experimental device applied about 800 lb to Schoeck during his run at 24.85 mph and effectively held him in a proper position for chest rotation onto the

steering-post pad. As is the case of the dashpot operation, the release of this pressure has not yet been developed in these tests. It may require the combination of a cylinder and a piston with a metering pin like that in the bumper. Once its design is determined, however, the device should be particularly effective with higher impact loads. This device would be placed under the seat and connected by tubing to the hydraulic bumper. It could operate on the ends of the seat belts alone, or the force could be applied to the seat structure and the belts. The present mobility of cylinder and piston systems could make possible a simple structure under and behind the seat that would meet the requirements of seat movement and necessary strength.

Figure 15-10 shows the test-cart deceleration and the seat-belt forces on Schoeck, weight 185 lb, at impact velocities of 19.45 and 24.85 mph. A padded steering post with 3-in. hydraulic absorption decelerated the upper torso and head without causing injury. A standard seat belt was used, but a seat-belt tightener, connected to the hydraulic bumper, held the man securely in position during the impact. In these tests no hydraulic relief or dashpot effect was obtained, although the inclusion of these factors would further reduce the deceleration forces on the man.

In Figure 15-10 there is an additional set of curves for a 40-mph

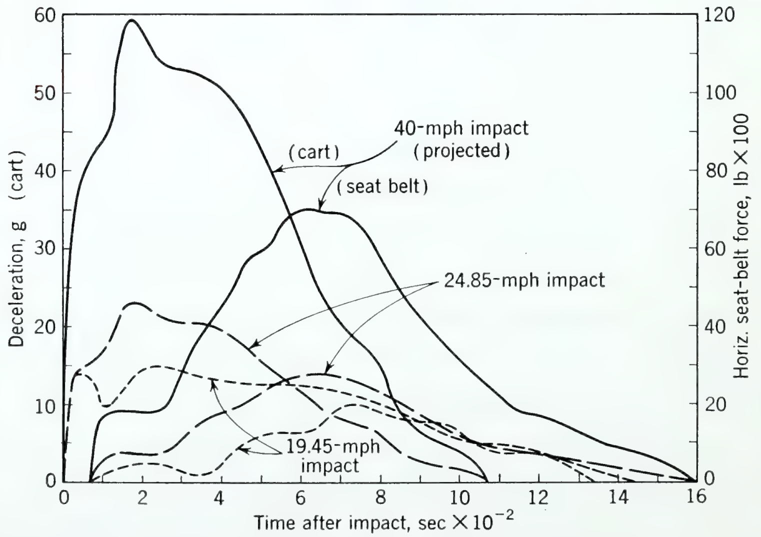


FIG. 15-10 *Cart deceleration and horizontal seat-belt forces measured on Schoeck (185 lb) at 19.45- and 24.85-mph impacts with a padded steering wheel and a post with a 3-in. absorption and a standard seat belt with automatic tightener—projected to 40-mph impact.*

impact that was calculated from the slower-speed test data shown. It is observed that with the hydraulic bumper the maximum deceleration would be about 60 g and the equivalent maximum horizontal seat-belt force would be about 7,000 lb. Although the maximum horizontal seat-belt force measured on a human being in these deceleration tests was 2,800 lb, it is believed that, with the proper hydraulic relief or dashpot effect, forces approaching 6,000 lb at a 40-mph impact velocity could be safely tolerated by the automotive passenger.

Figure 15-11 is a summary of the deceleration for the equivalent

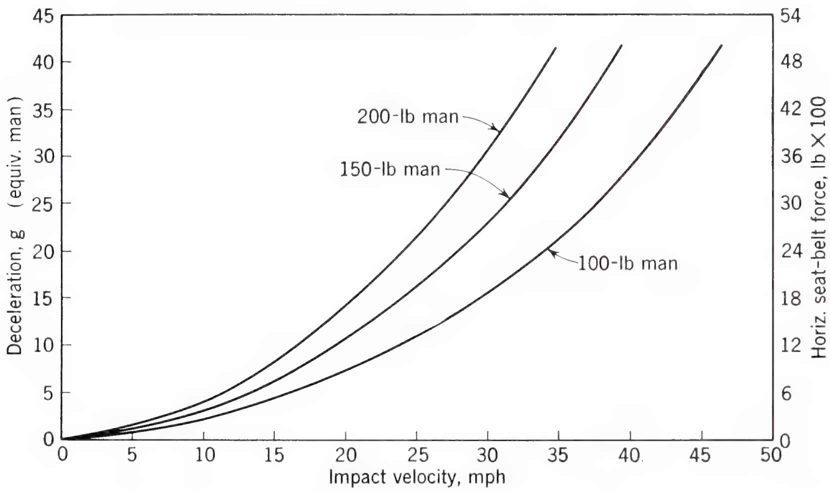


FIG. 15-11 *Hypothetical deceleration and horizontal seat-belt forces for different impact velocities and various passenger weights with fixed seats, a seat-belt tightener, hydraulic shock-absorbing bumpers, and 0.625 critical damping in series with the seat belt.*

man and of the maximum horizontal seat-belt forces for different impact velocities and different passenger weights, with fixed seats, a seat-belt tightener, hydraulic shock-absorbing bumpers, and critical damping in series with the seat belt. At 40-mph impact velocity, the theoretical horizontal force would be about 6,600 lb, which is about the same belt force as that found with dummies for the 28-mph impact of the 1949 Chevrolet (Figure 15-4). Thus, the comparison shown here of equivalent seat-belt forces for 40- and 28-mph impacts is an indication of the reduction in collision forces to be gained for the human automotive passenger. This ratio applied to equivalent cars would result in a force reduction of about one-half. The force of about 6,000 lb is believed to be within the human tolerance for minor injury.

ENGINEERED SEAT BELTS

The requirements for the reduction of the forces on the passenger held by a seat belt include the proper elasticity; the use of rigidly locked seats that do not move on impact; a seat-belt tightener, preferably a hydraulic piston mechanically connected to the belt ends and operated by the fluid pressure built up in the shock-absorbing bumper on impact; and a proper hydraulic relief or dashpot with metering fluid restriction that allows a 2- or 3-in. travel of the belt ends with critical damping and has a restraining force proportional to the velocity of movement.

Similar conditions must be realized for the free swing of the passenger beside the driver. Although no difference has been found between a free-swing situation and the forward deceleration of the rotating chest and head on the steering pad, impacts with any surface except the body itself may create injury. Therefore, the body must be held in proper position by the seat belts. In addition, adequate clearance for the body and head before the windshield must be given in order to preclude contact. The swinging arms as well as the lower legs may strike under the dash at the windshield, but proper padding will reduce injury, since the impacting masses of these extremities are relatively small.

Through proper mechanical construction of the driver's compartment, use of hydraulic shock-absorbing bumpers, engineering control of the positions of human bodies in seat belts, and application of proper decelerating techniques for the extremities, the total impact force transmitted to human bodies can be reduced to about 25 per cent of that transmitted on seat belts in present automobiles. This would create a reduction of one-half in annual injuries, deaths, permanent disabilities, and property-damage costs.

BIBLIOGRAPHY

1. Accident Facts, National Safety Council Annual Report for 1960, 1961 ed., p. 40, National Safety Council, Chicago.

2. Facts and Figures about Air Transportation, *Aviation Week,* vol. 74, no. 18, pp. 69–100, May 1, 1961.

3. Fredericks, R. H., and R. W. Connor: Crash Studies of Modern Cars with Unitized Structure, paper presented at SAE National West Coast Meeting, San Francisco, Aug. 16–19, 1960.

4. Minnesota Safety, Minnesota Safety Council Annual Report for 1960, 1960 ed., p. 4, Minnesota Safety Council, St. Paul.

5. Moore, John O.: A Report of the Testimony from Members of the Cornell Committee for Transportation Safety Research, Seat-belt Hearings

in the U.S. House of Representatives, Automotive Crash Injury Research of Cornell University, New York, Aug. 8, 1957.

6. Ryan, James J.: Human Crash Deceleration Tests on Seat-Belts, *Aerospace Med.*, vol. 33, no. 2, pp. 167–174, February, 1962.

7. ———: Mechanical Reduction of Impact Forces by Automotive Design, *Research Review* (quarterly supplement of *Traffic Safety*, published by the National Safety Council, Chicago), vol. 6, no. 2, pp. 3–14, June, 1962.

8. ———: Reduction in Crash Forces, *Proc. Fifth Stapp Automotive Crash and Field Demonstration Conference, University of Minnesota, Minneapolis, Sept. 14–16, 1961*, pp. 48–89, April, 1962.

9. ——— and E. R. Podnieks: Safety Devices for Ground Vehicles, Automotive Safety Research Project, University of Minnesota, Minneapolis, July 31, 1958, and Sept. 1, 1960.

10. Severy, D. M.: Automobile Collisions on Purpose, *Human Factors*, vol. 2, pp. 186–202, November, 1960.

11. ——— and J. H. Mathewson: Automobile Barrier Impacts, ser. I, II, Institute of Transportation and Traffic Engineering, University of California, Los Angeles, 1956.

12. Summary Report, Automotive Crash Injury Research of Cornell University, 1953–1961, Automotive Crash Injury Research of Cornell University, New York, Apr. 3, 1961.

Chapter 16

MEDICOLEGAL INVESTIGATION OF TRAFFIC FATALITIES

*Richard Ford**

THE PREVALENT PRACTICE IN THE INVESTIGATION OF TRAFFIC DEATHS IS TO ascribe the particular event to alcohol, speed, somnolence, or incompetence. The event itself is usually labeled "accident," and the law is called upon to establish criminal negligence and, subsequently, civil liability for damages. The basis for legal proceedings is frequently tenuous. For instance, in the multiple vehicular collision or in the one-car collision with a fixed object, usually no competent examination is made of the wreckage to discover whether or not purely mechanical failure is the complete explanation.

Comparable with the neglect of "auto autopsy" is the failure to perform human autopsy. Nevertheless, a thorough autopsy must be done upon the dead driver, or presumed driver, and upon the dead pedestrian in order to establish, if possible, whether a physical defect existed that could have precipitated or contributed or did precipitate or contribute to the disaster. The range of information that may be elicited by careful medicolegal examination, which includes the scrutiny of clothing, surface injury, and visceral and skeletal damage, the chemical study of blood and visceral samples, and the interpretation of collateral information from medical and nonmedical sources, can be clearly demonstrated by case summaries.

Sometimes the careful examination of the body of a dead driver may disclose nothing beyond the characteristics of ballistic impact against whatever objects were involved in the collision. If the deceleration of the vehicle is rapid and the driver is not restrained by a seat belt, as is usually the case, he becomes a projectile striking the steering wheel with his chest, the windshield with his head, the uppermost rim of the steering wheel with his throat, the lowermost rim of the wheel with his abdomen, or the instrument panel with his knees and legs. Thus, drivers often sustain surprisingly characteristic injuries. Glass cuts on

* Harvard Medical School, Boston, Mass.

the forehead and face with skull and brain injury, crushing of the larynx with internal hemorrhage and edema leading to asphyxia, fracture of all 10 ribs of both sides of the chest cage with cardiac, pulmonary, or aortic damage, and rupture of the intestines or tearing of the mesenteric attachment of the bowel all correspond to the several surfaces encountered on impact.

The examination of bodies and clothing alone may not provide all the information that is required about the crash. It may be easy to establish that one cadaver among several was the driver, or it may be impossible if the major impact was indirect and there occurred lateral displacement of the body without much forward motion. However, with the aid of collateral information such as the pattern of the rubber brake-pedal cap ground into the sole of the right shoe or the indentation of the head-lamp beam-deflector switch in the sole of the left shoe, it may be possible to indicate not only who was driving but also that the driver was probably awake at the time of major impact.

Subsequent to the careful examination of body and clothing is the chemical examination of the heart's blood and of visceral samples. Alcohol determinations are always necessary. Although the role of alcohol is frequently overemphasized, the presence of high levels of blood alcohol in the absence of other significant factors may have grave importance, as is illustrated by the following two cases:

Case 1. In broad daylight a taxicab driver with a load of passengers drove at an estimated speed of 75 mph into a trolley car, killing himself and his passengers and injuring a number of trolley riders. His blood alcohol was 0.48 g/100 ml, a level often found to be lethal of itself.

Case 2. The driver of a station wagon, alone in the vehicle, collided with a supporting column of an elevated railway system at 11 A.M. His blood alcohol was 0.47 g/100 ml.

Middle-range or low levels of blood alcohol are difficult to assess as causes of fatal or nonfatal traffic collisions. Even more difficult to evaluate are the effects of therapeutic doses of prescribed medications, barbiturates, antihistamines, and many other agents in such low doses as to be undetectable by blood analysis. These may, in susceptible individuals, produce pharmacologic side effects or excessive responses causing the individual to be totally incompetent for driving. Overdosage with medications has led to many traffic deaths. For example:

Case 3. A middle-aged man driving alone slumped upon the steering wheel. His car jumped the median strip and collided with two other vehicles. He was ejected and killed. In his clothing was found a generous supply of phenobarbital and secobarbital properly prescribed by his physician for hypertension. His blood contained a mixture of the two barbiturates at a level of 2.7 mg/100 ml, a reasonable explanation

for his suddenly falling asleep. His heart and kidneys showed the effects of prolonged high blood pressure.

Even more difficult to discover, in spite of complete anatomic and chemical studies, are the cases of sudden driver incapacitation due solely to physiological mechanisms. If the driver's collar were tight, syncope caused by a hypersensitive carotid sinus would lead to instantaneous loss of control upon turning the head. Idiopathic epilepsy is sometimes concealed, as in the following case, from the motor-vehicle licensing authority because the automobile is an indispensable tool for the driver's job:

Case 4. In morning traffic a man was observed "struggling with the steering wheel." He collided lightly with two cars ahead in succession, drove off the road, overturned, and died of asphyxiation due to a crushed larynx. In the course of listing the driver's personal property, the medical examiner's secretary came upon a daily account book listing in minute detail his expenditures. One entry stated: Dilantin $2.50. Also in the book were the name of a physician and his telephone number. The doctor was called and expressed grief at the loss of his patient and personal friend, but he denied knowledge of significant disease. On being prompted about the Dilantin entry, he admitted knowing about the decedent's epilepsy and said that 6 months previously, while riding as a passenger, he had occasion to seize the wheel when the decedent had a fit.

In the same category of obscure disease is the following case:

Case 5. A beer-truck driver, forty years of age, was making a routine run in a familiar area. With him was a helper, who was a stranger. In spite of his long familiarity with the route, the driver stopped at a service station to ask directions to the next town. Soon thereafter, as he proceeded down a hill toward a red traffic light, he failed to brake, struck two waiting cars, knocked them aside, ran through the intersection, and rammed into a telephone pole, crumpling the cab so that the doors had to be opened with pry bars. The helper was dazed but relatively unhurt. The driver was dead, with minor surface injuries and no anatomically demonstrable cause of death. It was thought that the cramped position of his body might have led to asphyxiation, but proof of this was lacking. Inquiry into his history disclosed that in the past he had frequently been rewarded for his labor in trundling kegs of beer in and out of bars and taverns by being given "boilermakers" (a shot of whisky with a beer chaser). As a result of this occupational hazard, he had in the course of time become a steady consumer of large amounts of alcohol. About a week before his death, he ceased to drink but continued to work. On the basis of his confusion as to his whereabouts and subsequently his complete loss of control of the truck (which

was mechanically in good working order), it is reasonable to conclude that this beer-truck driver was the victim of a sudden fatal seizure due to alcohol withdrawal. In court this diagnosis would be difficult to sustain, but it is medically plausible.

To add to the list of anatomically intangible but nonetheless human-failure causes of fatal accidents, the following two cases are cited:

Case 6. A car traveling at a moderate rate of speed in a downtown area swerved off the street, struck a mailbox against which two young men were leaning, and pinned them against a brick wall. Each man lost a leg, but both survived. The driver was badly hurt, in shock, but conscious. When questioned just prior to surgery, he stated that he either coughed or sneezed and consequently lost control. Because of incorrect diagnosis, his torn mesentery, a result of abdominal contact with the lower rim of the steering wheel, was not surgically repaired, and he died of hemoperitoneum, in spite of repeated transfusions.

Case 7. The author observed a car lying upside down in a field beyond a ditch. No person and no moving vehicle other than the observer's were in sight. In the car were a woman and child. When helped out, frightened but unhurt, she showed a bee sting on the back of her neck, a hard white wheal surrounded by a rim of reddening. The sting was clearly demonstrable because she survived. If, however, she had been stung while driving along the edge of a ravine, her sudden death might well have prevented any local reaction and the cause of death would have remained unknown. How many of us can unflinchingly endure a bee sting or a hot cigarette ash in the eye?

The car that is driven directly, without braking, into a tree, pole, or abutment may have been so guided with full intent of the driver. Two cases emphasize this point:

Case 8. On a limited-access road a car was observed as it crashed at a high rate of speed into the abutment of an overpass. Only one body was present in the debris. In the driver's right temple was a contact gunshot wound. In the wreck was found a .38-caliber revolver with one empty cartridge case. Depressed by pressures from wife and mother, the driver made suicide doubly sure.

Case 9. A vehicle crashed into a bridge abutment. Dead were a man and his infant daughter. A police check disclosed that shortly before the "accident," the man had surprised his wife with a paramour. He winged his rival with a gunshot, but the guilty couple fled. He wrote his spouse a hate note and told her of his plans. Into the car he put the child for murder and suicide.

In the nine cases cited above, anatomic manifestations of disease were absent or only of indirect import. There are many cases of traffic deaths in which the dead drivers show significant disease. Yet even

then the problem is not always solved. Sixty per cent of those who die suddenly of coronary artery disease show no abrupt change. The same degree of arterial narrowing was present a day, a week, or a month before death. Stress is the trigger that sets off the demand for increased heart action and a greater coronary arterial blood supply. When the arteries are too narrowed, the demand cannot be satisfied and the heart stops. Driving creates emotional and sometimes physical stress. In the case of an unexplained collision, the presence of severe coronary artery disease in the absence of demonstrable mechanical or environmental defect is at least presumptive evidence of cardiac disease as a precipitating cause of the crash. In other instances, dramatic disease changes clearly demonstrate the cause of the event:

Case 10. A railroad police officer was driving his lieutenant in slow-moving traffic in the heart of the city's shopping district. The car crashed into a parked vehicle. The officer was dead, the lieutenant shaken but unharmed. The driver's heart had an area of infarction 7 to 10 days old. He had made no complaint of pain and showed no outward sign of disability. However, to survive his disease, he should have been flat in bed and under expert medical care.

Case 11. A forty-eight-year-old woman, driving alone, cut across the center strip and slammed into the crash fence. She was dead when help arrived. On her upper chest were several bruises. She had died of a ruptured dissecting aneurysm of the aorta. Such a disease state may occur spontaneously or result from impact to the chest. Did she lose control because the aorta burst, or did the aorta burst on impact with the crash fence? The chest bruises tended to indicate impact while the blood pressure was still raised. One can presume, therefore, that the impact killed her. The car was never examined for mechanical defect.

Case 12. A man at work complained of feeling ill. Although a friend offered to drive him home, he insisted on driving himself. En route he killed a nine-year-old girl pedestrian and collided with three parked cars. Two hours later he died in coma. Autopsy showed no injury. He had lost control and later died as a result of the rupture of a congenital cerebral aneurysm about the size of a small potato. For 6 years this man had experienced blackouts, but he continued to renew his driver's license in spite of them.

Traffic deaths involving pedestrians closely parallel driver crash deaths in that the same physical and emotional deficiencies that disable the driver make the pedestrian vulnerable. In most jurisdictions the pedestrian has the right of way. The law maintains the principle that a driver should control his vehicle at all times so as to avoid injury to persons or property, although in some cases it is impossible to drive an automobile in such a manner that another vehicle or a pedestrian cannot be compromised.

Case 13. On a Sunday morning there was a steady flow of traffic on an overloaded superhighway. A young college student ran out from behind the stone base of a bridge, was struck by a car, and was carried on its hood. In a "panic stop" by the driver, the student was thrown upon a grass plot. Police came with an ambulance. The boy was questioned at the hospital and admitted a suicide attempt. In his room at the college was found a note listing his personal deficiencies and emotional problems and stating his intention to kill himself.

Case 14. At midnight on a second-class road, a man attempted to place himself in front of traffic. Six cars miraculously avoided him. When apprehended, he had upon his person a note expressing his intention of self-destruction.

In cases 13 and 14 both young men were unhappy, self-avowed homosexuals. Their bland disregard of the drivers' feelings is matched only by the suicides in subway stations and the dynamiting of passenger aircraft for insurance purposes.

In respect to alcohol, the same problems that exist for drivers exist for pedestrians. Every pedestrian death, except for small children, demands a blood-alcohol determination. In fairness to the driver, a dead pedestrian should be examined for cataracts, which when mature are easily visible in the cadaver. Auditory defects are not usually recognized at autopsy. However, information concerning sight and hearing may be obtained from eye and ear specialists if they were ever consulted by the pedestrian. Senility, in the technical sense of feebleness and brain atrophy, is a common cause of the aged's lurching in front of a passing vehicle. Autopsy can disclose not only the extent of injury but also the extent of physical wasting.

Brain atrophy is not limited to the aged. The following case demonstrates that anatomic study of the dead can mitigate blame for the living:

Case 15. On his way to school a nine-year-old boy walked across half of a wide avenue. On the far side a gigantic, bright orange-red trackless trolley was loading passengers. The trolley accelerated rapidly. The lad stepped in front of it and was crushed. Even on external examination the extraordinary crushing injuries were evident. The chest was split open. Truck-tire tread marks marched across the abdomen and one leg. There was no doubt as to the cause of death—massive crushing injury—nor as to the manner—"accidental." The first judgment about the event was that the driver of the trolley had been receiving fares, making change, and greeting passengers and that therefore he had never seen the boy. On the other hand, the street was free of traffic except for the orange-red trolley. Why did the boy step in front of it? Autopsy was decreed because of the unreasonable nature of the event. Just prior to this procedure, a physician stated that 3 years previously the boy had been one of the first reported "cures" of tuberculous meningitis by the then new antibiotic,

streptomycin. The doctor was concerned about the efficacy of the treatment. Autopsy revealed complete atrophy and scarring of the right occipital lobe of the brain, caused by a tuberculoma impinging upon the right posterior cerebral artery. The boy had less than half of normal vision, and nobody knew it! He never saw the orange-red trolley.

Consider, as a magistrate, the following two cases:

Case 16. In coma, a man in his early fifties was brought by police ambulance to a hospital. He had a lacerated scalp and a fractured pelvis. Skilled neurosurgeons explored his right temporal brain, evacuated a blood clot related to a skull fracture, and found a directly lacerated brain. During the ensuing month, his condition improved, and he was sent home. One month later he returned to the hospital with signs of intracranial disorder. Reexploration of the injured area and a biopsy produced the diagnosis of malignant glioma, an infiltrating "cancer" of the brain, present long before the injury in traffic. Within a few weeks the man was dead. At autopsy the effects of surgery were apparent. The immediate cause of death was an extensive infiltrating malignant brain tumor blocking the ventricular drainage. This man died of disease; yet he had, as a result of being struck by a motor vehicle, received a skull fracture and brain laceration immediately overlying his malignant disease. Reasonable medicolegal handling of the case calls for a certification of death due to disease, but civil aspects raise the question as to whether the death was hastened by the skull fracture, brain laceration, local hemorrhage, or surgical exploration and reexploration. The ultimate question is whether the presence of the brain tumor led the man to step unwittingly in front of the car.

Case 17. At 3 A.M. an emaciated, tottering man in his sixties walked into the side of a passing motorcar. Trace evidence by dust marks proved that his sole contact had been with the side of the car. Autopsy showed long-standing carcinoma of the pancreas with extensive spread of disease to the extent of "carcinomatosis peritonei." By hospital record he was known to have had cancer of the pancreas for 3 full years. It was inoperable when surgically explored. Was this pedestrian death inadvertent or intended?

"Hit-and-run" or "skip" deaths of pedestrians may be a social tragedy. If the wage earner is killed and no monetary redress is available, the family and the taxpayers suffer:

Case 18. By a country road in an alder thicket lay the body of a man. His hat and glasses were 15 ft up the traffic stream; his rubbers, shoes, and lunch pail were scattered 20 ft down the stream. The single, fatal injury was a gouge through the left side of the skull. The night before had been dark and rainy, the black-top road glistening and wet. His left shoe showed an impressive pattern of tire-tread squeezing the

upper section of the shoe. On the wrong side of the road for pedestrian travel, he had been clipped by a vehicle, and the driver never knew it.

Case 19. An old man was found in a city street. One leg was broken, the other bruised. His right hip was fractured, the bone driven in. In the overlying skin was a shard of glass. His trousers had no tear or cut; the hip pocket was shredded by glass. The clothing and body reeked of muscatel, and on the next day the suspect's car stank equally of the wine.

Case 20. A truck went east, and a car went west. A boy lay in the street behind them. On his left forearm was a zigzag pattern of a heavy truck tire. The trucker denied hitting the boy, but his truck was impounded. Photographs were taken of the child's arm with a ruler in the plane of injury. The right front tire of the suspect truck, out of six possible tires, had the only pattern that matched the forearm bruises. A photograph of the truck-tire tread, with a ruler present for comparison, could be matched on projection inch for inch with the marks on the youngster's forearm.

Drivers and pedestrians are not alone; passengers, too, suffer in traffic. If the door catches do not hold, out go the passengers into the roadway. If they are confined within the car, without seat belts, the hazards are still immense:

Case 21. A woman in her late eighties, the grandmother in the family, complained of nausea and vomiting. The family doctor gave some help but urged cardiac consultation. On the way to the cardiologist's office, the old lady sat comfortably in the back seat of the car. In a minor collision the vehicle stopped suddenly. The grandmother was dead. Postmortem showed a ruptured heart with a 10-day-old myocardial infarct. It is possible that this woman died only because of the excitement. However, an impact against the front-seat back could well have caused the rupture.

From these 21 summarized cases, it is apparent that the injunction "Seek and ye shall find" is applicable to the medicolegal investigation of traffic deaths. Autopsy alone can, on occasion, fully explain the event. More often, however, chemistry, past history, and medical and social sources of information must be explored in order to interpret the anatomic findings. Not to be forgotten is the combination of vehicle failure or deficiency and human failure or deficiency. A traffic fatality should not be dismissed with easy phrases such as "The driver fell asleep," "He lost control of his car," or "He admitted that he had been drinking." Clearly, no investigation of traffic death is adequate without broad medicolegal study of the driver and the pedestrian.

Part E

HUMAN FACTORS IN SENSORY SUPPLEMENTATION

Chapter 17

SENSORY SUPPLEMENTATION: AN INTRODUCTION

John K. Dupress and Thomas B. Sheridan†*

ONE OF THE MOST IMPORTANT PROBLEMS IN THE AREA OF EXTENDING MAN'S sensing abilities through the use of physical instrumentation is that of supplementing the senses available to the blind and the deaf-blind. It is estimated that vision alone accounts for 65 per cent of all the afferent neurons of man and hearing accounts for another 2 per cent.[1] The problem is to provide, through man-made instrumentation, the proper information to the remaining senses so that as much as possible of the lost function can be regained.

The human factors engineering problems of supplementing the lost visual senses may at first seem less exotic, more mundane than, say, providing for man in space. The needs appear straightforward enough—to permit the blind individual to read ordinary type and to prevent his tripping on curbs and colliding with obstacles as he walks. He is willing to read or to walk much more slowly and with greater restrictions than the sighted individual. He seemingly places no inordinate demands on his electromechanical aids in terms of miniaturization or reliability. However, at second look, the problems of engineering devices to aid the visually deprived are extremely complex and difficult. There exists no coherent body of theory by which to plan and predict, except that borrowed from other engineering areas, which may or may not be relevant. There are engineering development problems, to be sure, but more than this there is great dearth of fundamental knowledge as to (1) what information to select from the environment and (2) how to encode such information for best "plugging" it into the human nervous system.

When all man's sensory channels are functioning normally, the individual is flooded with excessive and redundant information. In order to prevent the human brain from being jammed, the body has built-in

* American Foundation for the Blind, New York, N.Y.

† Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, Mass.

[1] Warren S. McCulloch, Massachusetts Institute of Technology, personal communication, April, 1961.

protective mechanisms to limit the bandwidth of the end organs, to perform complex actions at local neural levels, to throw away information at various places in the central nervous system, and to establish priorities among senses to help the individual to select and accommodate.

When vision or both vision and hearing deteriorate, the human frequently has too little information for decision making. The remaining sensory channels must be utilized with maximum efficiency. Instruments must be designed to permit access to data ordinarily available only through vision (and hearing in the case of the deaf-blind).

Because the blind and deaf-blind must be so much more efficient in their use of available data, information coding is most critical. Since the channel capacity (information bits per second) is significantly less (by almost any measure) in the case of other sensory channels than it is in vision, we must maximize the usefulness or meaningfulness per bit of information. This will entail consideration not only of the second-by-second information rates of the peripheral sense organs but also of the human brain as a storage medium (10 to 100 billion bits) and its unique ability to retrieve programs selectively.

To help the blind person perform complex mobility or reading tasks, basic research is required to determine individual capabilities, training methods, and instrumentation design. Obstacle and step-down detection are indispensable to successful travel. The achievement of successful obstacle detection and location requires not only psychoacoustic research in depth but also the development of special laboratory equipment and measurement techniques and an awareness of certain testing problems that relate to sensory deprivation. Little is known about the transmission of navigation information through tactile maps, programmed muscle sensing, and mental constructs. When visual-print patterns are converted to braille symbols, research is required in skin sensation to determine the pertinent variables for this special kind of information acquisition. An infinite number of patterns normally recognized by vision must now be known by touch and muscle sensing.

Research being done on living creatures other than man that operate with low or no vision and/or no hearing should prove useful. Bats and porpoises are examples of sonic- or ultrasonic-ranging creatures. Much is known about the ultrasonic-ranging capabilities of the bat (see Webster's chapter, Chapter 25), and we are beginning to learn about the bat's nervous system. A relatively large amount of data is available on man's auditory system, but little is known about his sonic-ranging capability. Comparisons of man's highly developed nervous system (end organs and brain) with those of animals, birds, fish, and insects with limited and primitive visual and auditory systems may lead to some more theoretical foundation on which to base future sensory extensions and substitutes.

The main problems that arise from sensory deprivation in the form of blindness and the combination of blindness and impaired hearing are:

1. Direct access to the printed word and graphic forms. Up to the present time, blind or deaf-blind individuals have generally had no direct access to the printed word or graphic forms. Two instruments have been designed that have provided a few individuals such access. These are the optophone or Battelle Reader and the Visagraph (both of which techniques are further discussed in subsequent chapters by Freiberger and Murphy and by Abma). The former converts printed letters to monaural auditory signals, while the latter converts printed letters and graphic forms to raised lines. Otherwise printed material has been converted to the spoken word by human transcribers or to braille symbols that can be read tactually.

2. Direct access to the spoken word in the case of the deaf-blind. Some basic research has been done on selectively presenting elements of speech to deaf-blind individuals through touch. This work was done mostly at the Massachusetts Institute of Technology on the "Felix" project and at the University of Capetown, South Africa. Hopefully, research on speech analysis and speech synthesis could lead to the development of special instrumentation that would allow speech at the rate of a few words per minute to be converted into tactile information. The limitations of communication via the tactile and kinesthetic senses are gradually becoming manifest, as discussed in Chapter 24 by Bliss. For example, Geldard and his students have shown that no more than six or seven areas on the skin can serve reliably as independent and simultaneous vibrotactile channels.[2] However, certain deaf-blind individuals, by placing their fingers on the lips and throat of the speaker, are now able to perceive enough elements of speech for comprehension.

3. Mobility. During and since World War II, a great deal of research was done on radar, sonar, and infrared sensing systems. Much that has been learned about these active sensing systems for nonvisual obstacle detection can be applied, with appropriate "human engineering," to mobility devices for the blind (see the chapter by Benham, Chapter 22). However, existing prototype devices are far from a practical solution. At present the most satisfactory practical mobility device is the so-called "long cane"—essentially a crude extension of a single pointing, probing finger. However, even in the case of the simple cane, the effective sensory cues and their physical correlates are not well understood.

The form of coding and the particular transducers chosen, though they may represent the best design for particular kinds of environmental information, must not interfere with data about the environment that are

[2] F. A. Geldard, Adventures in Tactile Literacy, *Am. Psychologist*, vol. 12, pp. 115–124, 1957.

obtained directly by remaining sensory channels. Further, concerning any particular event, the data from the instrument must be presented simultaneously with data obtained directly by the senses. Sensory feedback loops that already exist (such as auditory feedback from footsteps or cane taps) must not be arbitrarily opened to couple with an instrument, for this may decrease any existing human capabilities. Failure to consider what the human being can already do or can be trained to do without auxiliary devices may result in a confusing jumble of directly obtained stimuli mixed with stimuli acquired by way of instrumentation.

Progress in developing sensory supplements or aids has been constrained both by a lack of funds for research and by a lack of proper interdisciplinary atmosphere. It was not until World War II that the Committee on Sensory Devices, National Research Council, specifically allocated funds for research in this area. There is at present little or no commercial profit incentive to induce private enterprise to enter this field, and government and foundation support remains limited. Until very recently, the development of sensory-aid instrumentation had been limited to solution of the electromechanical problems without adequate consideration of underlying behavioral problems and man-machine interactions. Participation by the medical profession had been limited to the problems of low residual vision and hearing.

However, we are told that the age of computers, servomechanisms, systems analysis, pattern recognition, and artificial intelligence is here. Engineers and psychologists interested in such ideas have come to see sensory extensions and supplements for the blind or deaf-blind as offering challenging and stimulating problems pertinent to their interests, needing practical solutions as soon as possible, and offering great personal satisfaction in return for their creative efforts.

Chapter 18

READING DEVICES FOR THE BLIND: AN OVERVIEW

Howard Freiberger and Eugene F. Murphy**

THE GRAVITY OF LOSS OF READING ABILITY DEPENDS ON THE RELATIVE importance of reading in a society, on the proportion of blind people in the group, and on the specific needs of an individual blind person for ready independent access to information stored in written or printed material, including charts, diagrams, maps, and photographs. Over the centuries problems have tended to increase. First, with an ever-increasing rate of literacy, the written word has assumed increasing importance in man's day-to-day life. Second, as our populations live to older ages, there is an increase in the proportion of visually handicapped [30]. Finally, as merely one example of blind persons in intellectual activities, including both the congenitally blind and those blinded later, there are now more who enter college than ever before. Their need to gain information from the printed page (including illustrations), already exceptionally great, will not diminish as they enter professional life. An excellent bibliography on many other aspects of blindness has been provided by Lende [35].

Many among the legally blind—whether congenitally or because of disease or injury—have some partial vision. Fortunately, some partially sighted persons have been enabled to read by direct vision supplemented by various aids [25]. Recent work has involved careful prescription of high-power spectacles [38], hand-held or desk-top magnifiers [13], or adaptations of closed-circuit television [73], supplemented by careful training with the prescribed devices. However, many among the partially sighted and all who are totally blind must rely on reading methods and developments such as those herein described.

* Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, New York, N.Y.

Parts of this chapter have appeared in essentially the same form in a previous article by the same authors, Reading Machines for the Blind, *IRE Trans. on Human Factors in Electronics,* vol. HFE-2, no. 1, pp. 8–19, March, 1961 [23], and are reproduced by permission of the IRE.

Reading systems for the blind, exclusive of magnifiers or other aids for the partially sighted, may be grouped into two broad categories, (1) systems requiring the intervention of sighted readers at one point or another and (2) systems allowing the blind independent access to the ordinary printed page.

Channels to the Blind Person

Within each of the categories mentioned above will be found existing systems using either the auditory or the tactual modality to substitute for the visual as the information channel to the user's brain. Numerous other systems for these senses seem likely to become practicable in the next few years.

Although experiments on electrical stimulation of the skin [32] or on phosphenes [80] have been tried and direct connection to the optic nerve or the visual cortex is occasionally proposed [8, 21, 75, 76], no practical system has been developed [62]. There are serious bioengineering problems, such as refining microelectrodes (now big compared with single nerve fibers), assembling arrays of electrodes, preventing gas formation at the electrode and ionization of the metal at the electrode tips (presumably prevented by use of alternating current or reversal of polarity), and physically connecting an extremely fine electrode to a specific nerve or bundle of nerve fibers without short-circuiting in a salty sea of body fluids. Biological problems, however, are probably more critical. Most surgeons are convinced that a foreign body cannot penetrate the outermost layer of the body indefinitely without infection or extrusion. Nevertheless, some extended experiments have been made by a few workers on the integral attachment of plastic artificial eyes through the conjunctiva [31, 55], insertion of plastic lenses through the cornea [67, 68], implanting of supports for dentures [36], and implanting of indwelling electrodes or catheters for periods of years in human beings [28] or in experimental animals [50].

Scarring at the connection between an electrode and a nerve must be minimized by substantially inert electrode material and by reducing mechanical, electrical, and chemical stresses. The connection might conceivably be attempted at the retina, the optic nerve, the higher optic radiation, or even the cortex of the brain. Electrical signals can be picked up from the relatively inaccessible visual cortex when lights are flashed before the eye, but, reciprocally, electrical stimulation may produce only such vague sensations as blue fog [52] or seeing stars; the complex combination of electrical and biochemical phenomena in a nerve is not necessarily reversible. Obviously there

are serious problems of coding external signals to produce those internal impulses normally handled by nerves or the brain itself. Probably other problems will be uncovered by partial solution of some of these. While one cannot now plug a television camera into the brain, as some suggest, future research may reveal reasonable possibilities for certain limited purposes, beginning with very simple patterns. Even separation of zones corresponding to left, ahead, and right might be useful in displaying the output of a guidance device, and braille has shown the great utility of a six-dot binary array. For the near future, however, the tactual and auditory channels seem best.

Reading Systems for the Blind

The relative acuities of the remaining senses will influence the choice among practical systems. With the deaf-blind, tactile methods obviously must be used. Others may be unable, in specific tasks, to use their hands for customary tactile reading methods; a device with audible output mounted on a typewriter, for example, might enable a blind person to monitor his touch typing [29]. With increasing age the auditory and tactual senses tend to deteriorate and the ability to learn complex tasks decreases, influencing the choice of reading method for the half of the blind population over sixty years of age. These multiple handicaps must be considered in the design of equipment for use by older blind persons. Personal preference, the nature of the reading task, titles available, and training background also have their influences on the system or systems that a person will use.

Successful practical systems of reading for the blind such as braille [34, 39, 80] and Talking Book [37, 80] generally require the intervention of sighted people, often in the preparation of material in special format such as braille books of embossed heavy paper or such as long-playing phonograph records. These methods are not discussed here, except incidentally as adjuncts to the systems described. Only systems, or constituent elements of them, allowing for the independent reading of the ordinary page are covered.

Past Experiments on Photoelectric Reading Machines

Since the early days of the photoelectric cell there have been efforts to design reading aids for the blind. Late in World War II and for some years afterward the Committee on Sensory Devices, with government funds, studied both guidance devices and reading machines. In "Blindness" [80], the Committee's final report, Cooper [11] summarized the many approaches tried by Radio Corporation of

America, Haskins Laboratories, and others; he concluded that a major limitation was lack of speechlike output. The fact that spoken words are natural, easily understood, audible analogs of their written versions explains why a spoken output seems so desirable in a reading machine for the blind.

By 1954 a number of organizations and individuals were working on automatic reading machines for business purposes, some psychologists and linguists had become concerned with speech synthesis related to the output problem, and agencies and individuals providing technical aids for the blind were beginning to feel that reading machines were not hopelessly far in the future. Unfortunately no one person knew all the others concerned, and no common meeting ground existed. Accordingly a series of technical sessions on reading machines for the blind [69, 70, 71] was arranged by the Veterans Administration. These sessions covered a wide range of background and historical information, reports on related developments, and suggestions for further work. The agenda formed a framework for classification of the many possible reading systems. An exposition of this classification system may be found in a paper [23] by the present authors. By the fourth such session [70], held August 23–24, 1956, the conferees reached substantial agreement on the characteristics desired in each of a series of mutually supplementary reading machines. There was general conviction that a number of practical devices could be developed based on the then current status of technology and psychology.

Current Programs

Soon after the August, 1956, session a number of proposals were developed and reviewed, leading to a series of projects under Veterans Administration sponsorship. Chapters 19 to 21 of this book were contributed by three of the scientific groups working on these projects.

Many other studies have been initiated over the years. St. Dunstan's, the English organization for the war-blinded, sponsored work immediately after World War II [7] and now has a scientific committee restudying the reading-machine problem. Der Deutsche Blindenverband, a German organization for the blind, is working with the center for information processing at the Technical College at Karlsruhe, already concerned with reading machines for post-office problems [33]. Work is also reported in Canada [2, 5] and the Soviet Union [54, 81]. Faculty members, students, and others have been meeting monthly at Massachusetts Institute of Technology for discussion of sensory research, continuing an interest stemming from the activities of the late Dr. Clifford Witcher [77, 78], a blind physicist formerly

at Haskins Laboratories, the American Foundation for the Blind, and later at the Massachusetts Institute of Technology. At this writing the American Foundation for the Blind is also conducting an international technological survey on aids for the blind.

READING-MACHINE DESIGN

The input end of reading machines for the blind, the section that transfers the information from the page into the machine, is closely allied to devices for automatic reading and character recognition. In the last decade, the military [44, 66], those studying machine translation between languages [3, 79], and large business activities concerned with automation [18, 24] have sponsored many developments in character recognition [65]. Though some business developments rely on specially designed type fonts, or magnetic inks, or both [10], many have elements potentially useful for the blind. The entire field was surveyed at the Optical Character Recognition Symposium in Washington, January 15–17, 1962.

Human Factors

Although there are similarities between reading-machine systems for business purposes and those for the blind, the presence of the human operator in the latter generally leads to a quite different system design and permits a more modest speed range. The ultimate receptor of the information being processed by reading machines for the blind, the human being, is endowed with cognitive skills that no machine has yet matched. The blind user can apply to his task reasoning of high order, sophisticated contextual cues, and skills in delicate manipulation. Thus the blind user potentially may substitute his human brain and his neuromuscular coordination for a complex section of the corresponding business machine. For example, a blind person, present during the scanning process, can do a great deal to orient the input copy, to select meaningful material, and to track along lines of print that generally are not spaced or leveled with a precision sufficient to make exclusively mechanical handling simple. The area scanned in detail may thus be decreased, and alignment methods may be simplified. In any event he can detect and correct many errors, resolve uncertainties, and bridge gaps that may occur in a machine output.

Depending on the specific reading tasks for which it is used, an acceptable machine for the blind may operate at speeds considerably

lower than those required for economical use in fully automated systems. Most blind people would probably be quite satisfied to read ink print at 180 words per minute (wpm) (perhaps comparable with talking books and with highly skilled use of braille). Many would tolerate reading at 90 wpm (slow public speaking), and for special tasks involving reading or mere identification even speeds of 15 to 20 wpm (slow Morse code) would be useful. Some subjects were reported [1] to feel that they would read for pleasure if they could attain 20 wpm. An acceptable reading speed depends, among other things, on the individual reader, the length, difficulty, and urgency of the material, the need for independent reading, and the degree to which the same information may be otherwise available to the blind reader. Ability to read independently, even if very slowly, such information as labels, brief instructions, recipes, or menus, or merely to orient letterheads correctly in a typewriter may have great value in personal and vocational life. Such capability for independent reading would be most valuable to the younger blind individual striving to advance but not yet senior enough in his organization to be allowed a secretary who could do much of this reading for him.

The mental and physical capabilities of human beings, including their ability to tolerate imperfections, may tend to simplify the design and production of reading machines for the blind. Fortunately those most likely to benefit from reading machines presumably are in the upper fractions of the blind population as distributed by intelligence, motivation, and literary ability.

Economic considerations of blind individuals, private philanthropies, and government agencies, however, impose difficult restraints to tax the designer's ingenuity. It is hard to estimate a reasonable price for a reading machine, including the cost of training in its use. If a machine provided the means to increase one's earnings, it would probably be bought at a price commensurate with such increase. A blind young physicist has commented that the independence created by even a limited reading machine would permit him to hold positions paying an additional $2,000 per year. Several hundred dollars seems to be a reasonable selling price for a home-type machine intended for fairly widespread individual use among the blind.

Portability also challenges the designer. An individually owned, home-type device should be portable, of the order of 5 to 10 lb, to allow for flexibility in operation. Installations intended for centralized use, as at schools or libraries, may be larger and much more expensive, however, in return for simpler, more rapid, highly dependable operation with minimal psychological stress or requirement for training. An intermediate-sized machine (for long-term use in a home or an office,

for example, and seldom moved) might cost $1,000, weigh 25 or 35 lb, and operate fairly rapidly with only occasional ambiguities in the output.

It is apparent that careful prescription from an armamentarium of systems will be needed. Great individual variations in human factors such as in audition or intelligence, in previous training, in vocation and avocation, and in economic and geographic situations preclude design of a single device ideally suited to all blind persons. Probably as with other types of prosthetic and sensory aids, prescription and checkout will be managed most wisely by teams of experts, and yet each specific phase (like training) will be conducted most skillfully by the expert in a particular field.

Examples of Devices Proposed or Actually Constructed

A number of systems, devices, components, and research studies illustrating the varied approaches to the problem will now be mentioned. The physically simpler examples will be described first, then the more complex ones.

Generally speaking, the psychological stress on the user is inversely related to the machine complexity. Reading a given line of type with a simple optical probe is almost impossible, and yet such function might be quite simple and rapid with one of the proposed complex-recognition devices.

A. *Simple Optical Probes.* Three similar optical-probe units, each having one photosensitive cell controlling the output, which are too rudimentary to be considered reading machines but which illustrate some principles of optical-to-audible transducers, are the Orientierungshilfe für Blinde [74], the Audivis [77], and the Hear-a-Lite [26]. A comparable single-cell device suggested by Moon [46] provides tactile output.

B. *Direct-translation Devices.* Examples of apparatus on the next level of sophistication, often called direct-translation machines, are Fournier d'Albe's early white-reading optophone [80], the improved black-reading version [4], several RCA developments [19, 47, 83, 84], Henke's optophone [29], and the Battelle aural reading device [1], which is based on the optophone principle and is covered in more detail by Abma in Chapter 19. In these systems the recognition process is performed by the user, who must learn to recover from the audible tone configurations the information originally presented in the letters on the page.

Direct-translation devices having tactile output are the Naumburg Visagraph [80], Snook apparatus [63], Sell device [58], Faximile visa-

graph [48], Rubbiani unit [70], and Surber device [72]. The visagraphs produce embossed replicas of the original print, the Snook, Rubbiani, and Surber devices employ an array of pins that are raised in a likeness of the pattern on the page, and Sell's device excites the skin electrically with a contact matrix charged with the pattern being read. Kallmann [69] suggested a hand-held device that could be moved over ink lines on a page, as in a diagram, or ink dots arranged as in braille [39]. In view of past experience with raised letters and charts, these direct-translation machines with tactile output seem relatively limited in speed but potentially useful for special tasks. This type is probably particularly valuable for graphical material and for the very small proportion of blind who are also deaf.

C. Intermediate or Integrating Device. When the Veterans Administration launched its research and development program on reading machines for the blind, it was realized that there were psychological problems and serious limitations in reading speed inherent in the simpler direct-translation systems but that the cost of the more complex type of recognition systems adapted from business devices was very high. Hans A. Mauch, noting the existence of rather well-developed plans and even some devices in both categories and mindful of their operational or cost disadvantages, postulated that a device intermediate between the two might provide a workable compromise. Instead of a conformal transformation of the spatial pattern visible on a page to an acoustical analog in pitch and time, Mauch [40] originally considered scanning an entire letter, integrating the information obtained at each of five levels, and responding to such information by producing a single complex sound, preferably like that of speech though perhaps like that of an unknown foreign language. This procedure avoids the tripartite sound pattern heard for most letters on direct-translation machines of the optophone type as the beginning, middle, and end of the letter are successively traversed. Use of a single-element complex sound per letter rather than a group of these chords presumably should allow for comprehension at up to three times the reading speed of an optophone. Thus without much increase in machine complexity Mauch contemplated an instrument potentially capable of considerably better performance.

Mauch now is working along somewhat different lines by using prerecorded fragments of actual human speech (the spelled speech [42, 43] described by Metfessel in Chapter 21) to be controlled in playback by a recognition front end [41], expecting that the cost and complexity can be kept at moderate levels by making use of the human operator. The recognition method used by Mauch involves an array of photosensitive elements so placed that they can most selectively respond to the black and white areas of the several letters in the font in question. Two spe-

cially shaped slabs of photosensitive materials are so placed as to detect the start of a letter and to control the instant when other carefully placed cells make their identifying snapshot of the letter image focused upon them. The cells' respective conductive or nonconductive states, through a relatively simple matrix, cause precise identification of the character.

By the use of appropriate interchangeable light-pipe masks later to be applied over the photocell array, Mauch hopes to accommodate the variations of many common fonts and yet to use one basic photosensitive arrangement. These masks, in a sense, will perform transformations of the single sensitive array into variants better able to discriminate between the letter shapes peculiar to different fonts, or vice versa.

D. Recognition Systems. Recognition-reading systems must first recognize the letter (or other character) on the page as one of the alphabet set being used. They may then sound the successive letters, usually by their names (to produce a spelling-bee effect) but preferably by spelled speech or, if feasible, in entire words. Such machines, though inherently quite complex physically, have promise of requiring the least learning and effort on the part of the user. Systems of this general type have been described by Schutkowski [57], Sharples [59], Davis and Hinton [12], and Flory and Pike [20]. Teletypesetter tape, salvaged from the printing industry, allows far easier recognition of individual characters than is possible with the finished printed document.

Haskins Laboratories, as related by Cooper in Chapter 20, has designed and nearly completed a recorded-dictionary machine capable of identifying individual letters from Teletypesetter tape, recognizing completion of each word, and then locating and pronouncing any of some 6,000 prerecorded words. This machine attempts to improve on the letter-pronouncing machines suggested above. Haskins Laboratories is also studying methods for synthesizing speech. These, in turn, may possibly allow construction of simpler machines with speechlike qualities.

Recognition front ends can also control tactile-output systems necessary for the deaf-blind. Examples of such systems are the IBM Braille Reader [9] and the Zuk Braille Book Reader [82]. The Friedrich device, schematized in Figure 18-1, for translating Teletypesetter tape to paper tape perforated in Grade I braille (fully spelled) is potentially able to transform a byproduct of the printing industry into a tape usable with the IBM reader. IBM [16] has developed a computer program capable of converting fully spelled words to the contracted Grade II braille, and Nemeth [49] is working on a similar program.

There is considerable current interest in the further development and use of systems for recognition of characters such as the numerals on cash-register tapes or other business documents and the addresses on envelopes. There is also work on pattern recognition in general. Numerous

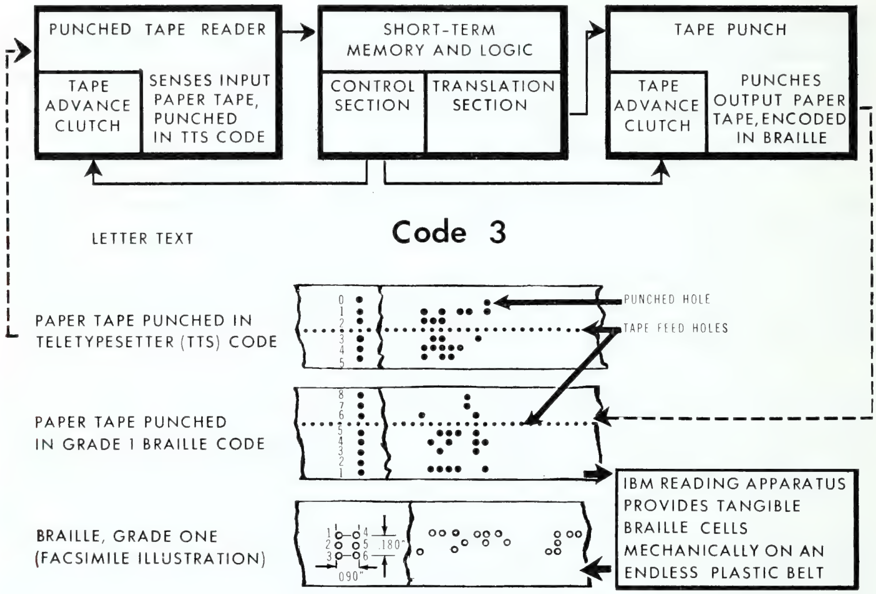


FIG. 18-1 *Block diagram of Teletypesetter-to-braille translator. The three principal sections of the device are shown, together with an adjunct unit, the IBM Reader [9], which provides a tangible braille-cell output in a moving belt from the braille-coded punched-tape input. The meanings of the code symbols in the TTS and braille encodings of Code 3 are tabulated below:*

TTS tape	**Braille (or braille tape)**
shift	*capital sign*
c	*c*
unshift	*o*
o	*d*
d	*e*
e	*space (only in tape)*
space	*number sign*
3	*c*

It is evident that the translation of even the simple expression Code 3 is not a one-to-one, character-to-character process. Producing the small o of Code in TTS requires two symbols, the unshift and the o, whereas in braille only one is needed, o. Reciprocally, producing the 3 in TTS requires only one symbol, 3, whereas in braille two are needed, number sign followed by c. (The digits 1 to 0 in braille are expressed by the number sign followed by the first 10 letters of the alphabet. Note that zero is the number sign followed by j, not the letters a or o.)

references may be found in a bibliography on artificial intelligence by Minsky [45] and in a "state of the art" report by Stevens [65]. A bibliography compiled by one of the authors [22] lists many items not mentioned here.

Work at Bell Telephone Laboratories [6, 14, 15] deals with the recognition of handwriting. Devices demonstrated have read handwritten

arabic numerals and the numerals 1, 2, etc., spelled out in longhand. The arabic numerals must be formed in an area including two poles from which radiate dividers. The pattern of crossing the dividers is used for numeral identification. The handwritten words must occur only within a selected area, the letters being of adequate size to cover properly the horizontal sensing zones used with this apparatus. The IBM Experimental Constrained Handlettering Reader, a similar device that reads hand-formed arabic numerals made with certain restraints, was demonstrated at the 1961 IRE International Convention in New York. Despite the restrictions mentioned, one still has reasonable freedom in positioning and shaping his writing for successful recognition by these systems.

The task is simpler if a business firm can control the printing of the characters and thus their type font and locations on business documents [10]. Shepard and his associates [60, 61] studied numeric and alphanumeric character recognition, including the development of a special set of characters readable by eye but intended for accurate recognition by machine [27]. Optical-recognition systems built on these principles are currently in commercial use, accomplishing such tasks as reading the serial numbers on canceled traveler's checks, vouchers for purchases by credit cards, etc. The Control Instrument Division of Burroughs has also built machines for reading serial numbers. Solartron Electronic Business Machines Ltd. [17] has been doing related work in England, and there are some units in commercial use. A report by Stone [66] describes an alphanumeric reader developed at Rome Air Development Center. Sprick [64], in Germany, is attempting character recognition by analysis of the shape of the outline. A recent patent issued to Rabinow [53] includes a mechanism for systematic vertical scanning or jitter, to improve on earlier stencil-matching recognition systems. IBM [51] has developed a character-recognition system for providing input to some of its computers.

Rosenblatt [56] has reported his work with Perceptron, a complex adaptive-network system that can learn to recognize patterns that come into its field of view. Presumably such devices could be taught to recognize any of a wide variety of graphic patterns. A versatile character-recognition system based on such self-adaptive methods or improvements may be a device far in the future, but, whenever achieved, it will be a powerful instrument hurdling with almost human ease problems of different type fonts, poor orientation or alignment, and imperfect characters.

The human factors of adapting any reading machine (from the simplest devices to complex self-adaptive machines) for practical problems of blind users are likely to be at least as difficult to handle as the solution of purely technical questions. Nevertheless, the potential value to human beings challenges all concerned to serious, prolonged, and realistic efforts.

BIBLIOGRAPHY

1. Abma, John S., et al.: Summary Report on the Further Development and Evaluation of Aural Reading Devices for the Blind, Battelle Memorial Institute, Columbus, Ohio, Veterans Administration Contract V1005M–1961, ASTIA AD-242 223, June, 1960.

2. Argyle, P. E. (Developed a reading machine for the blind at Royal Oak, B.C., Canada; tape recordings of output at American Foundation for the Blind, New York, and in authors' files.)

3. Baird-Atomic, Inc.: Russian Reader Uses Defocused Optics, *Electronic Design,* July 6, 1960. [Also in *Electromech. Design,* October, 1960, p. 56.]

4. Barr, Archibald, William Stroud, and Edmund Edward Fournier d'Albe: Optophone, U.S. patent 1,350,954, Aug. 24, 1920.

5. Beddoes, M. P.: An Audible Code, Department of Electrical Engineering, University of British Columbia, Vancouver, B.C., Canada, May, 1960. (Limited circulation; 10 pp. plus 5 figs. on audible code for a reading machine for the blind.) [See also M. P. Beddoes, E. Belyea, and W. C. Gibson, A Reading Machine for the Blind, *Nature,* vol. 190, no. 4779, pp. 874–875, June 3, 1961.]

6. Bell Telephone Laboratories: Recognition System Reads Ten Words, *Electronics,* vol. 32, no. 47, p. 100, Nov. 20, 1959. [Similarly reported in *Automatic Control,* vol. 11, no. 5, p. 22, November, 1959.]

7. Beurle, R. L.: Electronic Aids for Blind People, London. (An account of investigations carried out under St. Dunstan's Sensory Devices Committee during 1946–1951.)

8. Bosch, F. J. G. van den: Extending Vision with Electronics, Centre pour Electro-Microscopie, Antwerp, Belgium, ca. 1960.

9. Bryce, James W., and John N. Wheeler: Reading Apparatus (IBM Braille Belt Reader), U.S. patent 2,521,338, Sept. 5, 1950.

10. The Common Machine Language for Mechanized Check Handling, *Bank Management Publ.* 147, Bank Management Commission, American Bankers Association, New York, April, 1959.

11. Cooper, Franklin S.: Research on Reading Machines for the Blind, in Paul A. Zahl (ed.), "Blindness, Modern Approaches to the Unseen Environment," chap. 32, Princeton University Press, Princeton, N.J., 1950.

12. Davis, Edwin L., and Olar T. Hinton: Apparatus for Reading of Facsimile and Printed Matter by the Blind, U.S. patent 2,500,630, Mar. 14, 1950.

13. Davis, Jefferson P., Jr.: Results of Research Project on Development of Optical Aids for the Partially Sighted, University of Rochester, Rochester, N.Y., Veterans Administration Contract V1001M-1681, 1952.

14. Dimond, T. L.: Devices for Reading Handwritten Characters, *Proc. EJCC,* 1958, pp. 232–237.

15. ———: Experimental Device for Reading Handwritten Numbers, *Electronic Equipment,* vol. 6, no. 1, pp. 6, 7, January, 1958.

16. English to Braille Translation: How the IBM Computer Program

Works, IBM Data Processing Division, International Business Machines Corporation, White Plains, N.Y., Apr. 21, 1959.

17. ERA, Electronic Reading Automaton, Technical Director, Solartron Electronic Business Machines Ltd., Tower Hill, Dorking, Surrey, England, no date.

18. ERMA, Electronic Bookkeeper, *Research Ind. News Bull.*, vol. 7, no. 9, October, 1955.

19. Flory, Leslie E.: Reading Aid for the Blind, U.S. patent 2,457,456, Dec. 28, 1948.

20. ——— and Winthrop S. Pike: Apparatus for Indicia Recognition, U.S. patent 2,615,992, Oct. 28, 1952. [See also *Electronics*, vol. 22, pp. 80–86, June, 1949.]

21. Forman, Jan: Some Notes Concerning Possible Electronic Aids to Vision, July 5, 1960, and Further Notes on Electronic Aids to Vision, Sept. 9, 1960, unpublished memoranda, Arnoux Corporation, Los Angeles.

22. Freiberger, Howard: A Partial Listing of Selected References on Technical Aids for the Blind Appearing after January 1, 1950, Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, New York, Oct. 1, 1957. (With a series of later supplements.)

23. ——— and Eugene F. Murphy: Reading Machines for the Blind, *IRE Trans. on Human Factors in Electronics*, vol. HFE-2, no. 1, pp. 8–19, March, 1961.

24. Glauberman, M. H.: Character Recognition for Business Machines, *Electronics*, vol. 29, no. 2, pp. 132–136, February, 1956.

25. Gordon, Dan M., and Charles G. Ritter: Low Vision Aids in Ophthalmology, *N.Y. State J. Med.*, vol. 57, no. 21, pp. 3466–3470, Nov. 1, 1957.

26. Hear-a-Lite, developed at Bell Aircraft Corporation, Buffalo, N.Y., ca. 1956. (Information may be available from Batavia Lions Foundation, Inc., Batavia, N.Y.)

27. Heasly, Clyde C., Jr.: What Is Selfcheck®? Farrington Electronics, Inc., CCH-0029-1/60, 1960. [This is an adaptation of a paper, Selfcheck®, a New Common Language, read by C. C. Heasly, Jr., at the Western Joint Computer Conference, San Francisco, May, 1958.]

28. Heath, R. G. (ed.): "Studies in Schizophrenia," Harvard University Press, Cambridge, Mass., 1955.

29. Henke, Burton, Professor of Physics, Pomona College, Claremont, Calif., private correspondence.

30. Hurlin, Ralph G.: Estimated Prevalence of Blindness in the United States, American Foundation for the Blind, New York, October, 1953.

31. Jardon, Fritz W.: Anatomical Replacement Means, U.S. patent 2,688,139, Sept. 7, 1954.

32. Jerome, E. A., and H. Proshansky: Electrical Stimulation, Appendix P in Research on Guidance Devices and Reading Machines for the Blind, A Final Report of Work Done at Haskins Laboratories, New York, Dec. 31, 1947. [See also chapter by same authors in Zahl [80].]

33. Kazmierczak, H.: The Potential Field as an Aid for Character Recognition, UNESCO/NS/ICIP/H. O. 101, June 19, 1959.

34. Krebs, Bernard M.: Transcribers' Guide to Standard English Braille, New York Guild for the Jewish Blind, New York, 1953.

35. Lende, Helga: Books about the Blind—A Bibliographical Guide to Literature Relating to the Blind, American Foundation for the Blind, New York, 1953.

36. Lew, Isaih: Progress in Implant Dentistry—An Evaluation, *J. Am. Dental Assoc.*, vol. 59, pp. 478–492, September, 1959.

37. Library Service for the Blind, American Foundation for the Blind, New York, June, 1954.

38. Linksz, Arthur: Visual Aids for the Partially Sighted, *Am. J. Ophthalmol.*, vol. 41, no. 1, pp. 30–41, January, 1956. [Also appears in *Guildcraft*, vol. 31, no. 6, pp. 15–26, June, 1957.]

39. Lipton, A.: Introductory Notes on the Braille System of Reading and Writing for the Blind, Prosthetic and Sensory Aids Service, Veterans Administration, New York, April, 1959.

40. Mauch, Hans A.: Reading Machine, U.S. patent 2,907,833, Oct. 6, 1959.

41. ———: Summary Report on the Development of a Reading Machine for the Blind, Mauch Laboratories, Dayton, Ohio, Veterans Administration Contract V1005M-1943, June 30, 1960.

42. Metfessel, Milton: Summary of Developments in Veterans Administration Project V1005M-458 with the University of Southern California (June, 1955 through June, 1960), Research and Development Division, Veterans Administration, New York. (A 15-page report on spelled speech.)

43. ——— and Constance Lovell: Spelled-Speech as Output for an Automatic Reader, Communications Laboratory, Psychology Department, University of Southern California, Summary Report to Veterans Administration, Prosthetic and Sensory Aids Service, June 9, 1955–June 30, 1960, Contract V1005M-458, January, 1961.

44. Minot, Otis N.: Automatic Devices for Recognition of Visible Two-dimensional Patterns: A Survey of the Field, *USN Electronics Lab. Tech. Memo.* 364, June 25, 1959. (This is an unpublished memorandum.)

45. Minsky, Marvin: A Selected Descriptor-indexed Bibliography to the Literature on Artificial Intelligence, *IRE Trans. on Human Factors in Electronics,* vol. HFE-2, no. 1, pp. 39–55, March, 1961.

46. Moon, Robert J., Associate Professor of Physics, University of Chicago, private correspondence.

47. Morton, George A., and Leslie E. Flory: Reading Aid for the Blind, U.S. patent 2,420,716, May 20, 1947.

48. National Research Council, Committee on Sensory Devices: Instruction Manual for the Faximile Visagraph, Radio Inventions, Inc., New York, Contract VAM-21223, Subcontract 19, June 30, 1947. (Limited availability.)

49. Nemeth, Abraham, reporting at the Conference on Automatic Data Processing and the Various Braille Codes, Mar. 17–18, 1961, at Massachusetts Institute of Technology. (Minutes available from American Foundation for the Blind, New York.)

50. Olds, James: A Preliminary Mapping of Electrical Reinforcing Effects in the Rat Brain, *J. Comp. Physiol. Psychol.*, vol. 49, no. 3, pp. 281–285, June, 1956.

51. Optical Reader Bridges the Gap between Printed Page and Computer, *Automatic Control,* vol. 13, no. 5, pp. 54, 55, November, 1960. [Also in *J. Franklin Inst.*, vol. 270, no. 6, pp. 547–548, December, 1960.]

52. Penfield, Wilder, and Theodore Rasmussen: "The Cerebral Cortex of Man," The Macmillan Company, New York, 1950.

53. Rabinow, Jacob: Reading Machine, U.S. patent 2,933,246, Apr. 19, 1960.

54. Reading Machine for the Blind, *Les Aveugles vous parlent,* vol. 4, no. 35, pp. 1–2, 1958. (An illustrated account in French of a machine developed in the U.S.S.R. A later version of this optophone-type machine was demonstrated during a Russian exhibition at Earl's Court, London, July, 1961.)

55. Rolf, David E., Hudson D. Fowler, Jr., and Thomas S. Gerspacher: Artificial Eye Implant, U.S. patent 2,516,804, July 25, 1950.

56. Rosenblatt, Frank: The Design of an Intelligent Automaton, *ONR Research Rev.* NAVEXOS P-510, pp. 5–13, October, 1958.

57. Schutkowski, Georg: Optisch-elektrische Einrichtung für Blinde zum Umwandeln von Schriftzeichen in Sprechlaute oder andere von Blinden wahrnehmbare Zeichen, z. B. Blindenschriftzeichen, German patent 570,403, Feb. 15, 1933.

58. Sell, William O.: Aiding Device for Blind Persons, U.S. patent 2,327,222, Aug. 17, 1943.

59. Sharples, Alfred R.: Audible Reading Apparatus, U.S. patent 2,228,782, Jan. 14, 1941.

60. Shepard, D. H., P. F. Bargh, and C. C. Heasly, Jr.: A Reliable Character Sensing System for Documents Prepared on Conventional Business Devices, *1957 IRE Conv. Record,* pt. 4, pp. 111–120.

61. ——— and C. C. Heasly, Jr.: Photoelectric Reading Feeds Business Machine, *Electronics,* vol. 28, no. 5, pp. 134–138, 1955.

62. Shipley, Thorne: Problems of Corticogenic Vision: A False Hope for the Blind, *Am. J. Ophthalmol.*, vol. 47, no. 3, pp. 358–363, 1959.

63. Snook, Homer C.: Apparatus to Enable the Blind to Read, U.S. patent 1,889,576, Nov. 29, 1932.

64. Sprick, Walter: Apparatus for Identifying Line Traces, U.S. patent 2,738,499, Mar. 13, 1956. (Relates to a device for identifying printed or handwritten symbols.) See also Walter Sprick, Character Reader, U.S. patent 2,838,602, June 10, 1958. (Relates to an apparatus for reading printed matter and handwritten characters.)

65. Stevens, Mary Elizabeth: Automatic Character Recognition, A State-of-the-Art Report, *Natl. Bur. Standards (U.S.) Tech. Note* 112, May, 1961. (For sale by the Office of Technical Services, U.S. Department of Commerce, PB 161613.)

66. Stone, W. P.: Alpha-numeric Character Reader, *USAF RADC Tech. Note* 56-194, ASTIA AD-103 237, June, 1956.

67. Stone, William, Jr.: Study of the Patency of Openings in Corneas

Anterior to Interlamellar Plastic Artificial Discs, *Am. J. Ophthalmol.,* vol. 39, no. 2, pt. 2, pp. 185–196, February, 1955.

68. ——— and Elizabeth Herbert: Experimental Study of Plastic Material as Replacement for the Cornea: A Preliminary Report, *Am. J. Ophthalmol.,* vol. 36, no. 6, pt. 2, pp. 168–173, June, 1953.

69. Summary, Third Technical Session on Reading Machines for the Blind, Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, New York, February, 1956.

70. Summary, Fourth Technical Session on Reading Machines for the Blind, Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, New York, Mar. 27, 1957.

71. Summary, Fifth Technical Session on Reading Machines for the Blind, Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, New York, Dec. 31, 1959.

72. Surber, Curtis M.: Reading and Writing Device for the Blind, U.S. patent 2,866,279, Dec. 30, 1958.

73. Telereader, Industrial Television Systems, Tampa, Fla., ca. 1960.

74. Ukovic, Gregor: Orientierungshilfe für Blinde, German patent 717,223, Jan. 22, 1942. (A device wherein the intensity of incident light is used to vary the frequency of a tone in an earpiece.)

75. Using Electronics to Help Blind, *Electronics,* Dec. 1, 1957, p. 24.

76. Volpe, Ignatius: Sight through Electronics, folder sponsored by Hempstead Lions Club, Hempstead, N.Y., no date.

77. Witcher, Clifford M.: The Optical Probe—A New Tool for the Blind, *Technol. Rev.,* vol. 59, no. 2, December, 1956.

78. ———: Some Communication Aspects of Visual Prosthesis, paper presented at meeting of the AAAS, Boston, Dec. 28, 1953.

79. Yovits, M. C.: Navy Machine Translation Research, *ONR Research Rev.* NAVEXOS P-510, pp. 8–11, August, 1960.

80. Zahl, Paul A. (ed.): "Blindness, Modern Approaches to the Unseen Environment," Princeton University Press, Princeton, N.J., 1950.

81. Zemtsova, M. I., and I. A. Sokolyanskii (eds.): Osobennosti Vospriyatiya u Slepykh i Kompensatsiya Otsutstvuyushchego Zreniya pri Pomoshchi Tiflotekhniki [Particulars of Perception by the Blind and Compensation for the Absence of Vision by the Aid of Typhlotechnology], *Proc. Inst. Defectol.,* in *Bull. Acad. Pedagogical Sciences,* issue 90, 1957. (In Russian.)

82. Zuk, Gerald H.: Braille Book Reader, U.S. patent 2,891,324, June 23, 1959.

83. Zworykin, Vladimir K., and Leslie E. Flory: Electronic Reading Aid for the Blind, U.S. patent 2,457,099, Dec. 21, 1948. [See also *Proc. Am. Phil. Soc.,* vol. 91, no. 2, pp. 139–142, April, 1947, and *Electronics,* August, 1946.]

84. ——— and James Hillier: Optophone, U.S. patent 2,451,014, Oct. 12, 1948.

Chapter 19

THE BATTELLE AURAL READING DEVICE FOR THE BLIND

*John S. Abma**

BATTELLE'S RESEARCH ON A READING DEVICE FOR THE BLIND HAS BEEN directed toward the development of a relatively simple and inexpensive device. It was recognized by the Veterans Administration, sponsors of the research, that a low-cost instrument could probably not produce a code that would be ideal from the user's viewpoint; that is, it could not provide the user with spoken English, spelled speech, or some other immediately useful output. Instead, a code might be provided that would require extensive training before it could be understood, and low reading speeds might have to be accepted. Nevertheless, if a simple device could be shown to have some usefulness, then it could be made available to large numbers of blind people without delay.

The first decisions needed were in the area of aural-code selection. The desire for a simple device led to a class of codes described as direct-translation aural codes. These codes exhibit a close correspondence between letter shapes and output sounds. Usually spectrographs of the codes reveal the original printed letters, although they may be distorted. Past attempts to develop direct-translation readers suggested a number of codes that might be suitable. Those which appeared most promising were the code of Fournier d'Albe's optophone [1], first demonstrated around 1914, and the sweep code of a Radio Corporation of America device developed in the 1940s—the RCA A-2 reader [4]. In both these devices, a thin vertical-scanning slit is moved across the print from left to right. Only a small portion of a given letter is sensed at any instant: the total letter pattern is integrated over time by the listener. Also, in both devices, tones of low pitch represent low parts of a letter, and higher tones represent higher letter segments. Another feature in common is the sequencing of tones heard. In reading a letter from left to right, the

* Systems Engineering Division, Battelle Memorial Institute, Columbus, Ohio.
The research reported in this chapter was performed under Contract V1005M-1261 with the Veterans Administration.

first sounds to be heard represent the left edge of a letter, and the last sounds correspond to the right edge.

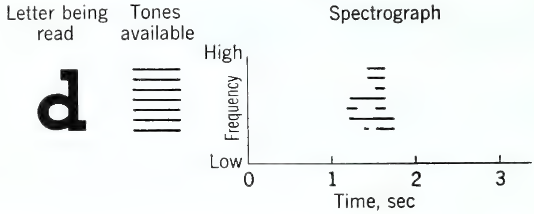


FIG. 19-1 *Construction of a hypothetical spectrograph for a seven-channel optophone code.*

The optophone and sweep codes differ markedly in the means employed for sensing print. The optophone codes employ multiple sensitive zones, each zone responding to a different "level," or horizontal slice, of the letter. A different sound frequency is assigned to each zone, or channel. An optophone-code spectrograph for the letter d is diagramed in Figure 19-1. A seven-channel code is illustrated. The reading speed shown is one letter per second, or 12 wpm.

The sweep codes employ a single sensitive zone that is moved rapidly up and down through the thin vertical-scanning aperture. The rate of vertical motion may be as high as 30 cps. Any attempt to increase this sweep frequency above 30 cps would probably bring with it two disadvantages: appearance of the sweep frequency as a background tone in the audio output and loss of pitch character for the lower tones.

Output signals are produced whenever print is detected during vertical-sweeping excursions of the sensitive area. If print is detected at the top of the aperture, a high tone is heard; if at the bottom, a low tone. These tones, high or low, are pulsed at the sweep frequency. Figure 19-2 illustrates the sweep pattern for lower-case d.

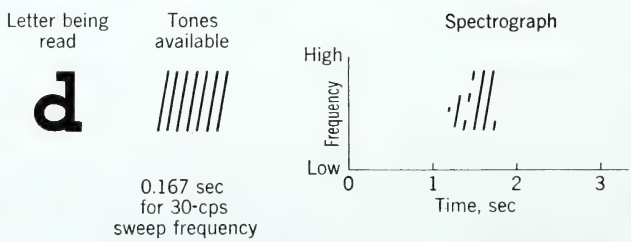


FIG. 19-2 *Construction of a hypothetical spectrograph for a sweep code.*

For clarity, this diagram exaggerates the time elapsing between vertical sweeps—there would actually be 30 instead of 10 sweeps for the reading of d at 12 wpm. It should also be noted that no tones are illustrated

for the return of the sensitive spot from the top to the bottom of the aperture. This return blanking was not a feature of the RCA A-2 reader, but was suggested by a committee investigating direct-translating home readers [3]. To reduce the complexity of the output sound patterns, the same committee also recommended a reduction in the number of tones available in the sweep code.

One inherent advantage of a sweep code is its tolerance of tracking error. That is, slight deviations of the scanning aperture above or below the center line of print being read would produce practically no change in the sound patterns heard. All the tones would merely be transposed up or down a few cycles, depending upon the degree of misalignment.

With the optophone code, however, some relatively gross changes in sound pattern might result from a slight misalignment or tracking error. Since the Battelle device was intended for manual print tracking, error tolerance was an important consideration. The d'Albe optophone employed mechanical print tracking; so no information on a manually tracked optophone was available. However, graphic analyses indicated that some improvement in tracking-error tolerance could be obtained by increasing the number of channels in an optophone code.

Battelle studied both a sweep code incorporating the above features and an eight-tone optophone code. The tones available for pattern representation in each case are shown in Figure 19-3.

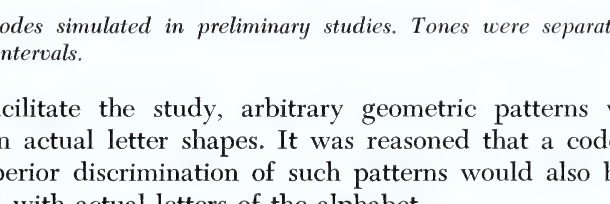


FIG. 19-3 *Codes simulated in preliminary studies. Tones were separated by equal logarithmic intervals.*

To facilitate the study, arbitrary geometric patterns were read, rather than actual letter shapes. It was reasoned that a code that permitted superior discrimination of such patterns would also be superior when used with actual letters of the alphabet.

The optophone code gave significantly more correct pattern identifications than did the sweep code. On the basis of this evidence and in the light of engineering-feasibility estimates, it was decided to adopt an optophone code. The problem of tracking deviations would be controlled by increasing the number of channels.

Further experiments were conducted to determine optimum values for other code parameters. A logarithmic intertonal relation was found

to be superior to a harmonic progression, and sine waves resulted in less masking of upper tones by lower tones than was the case for various complex wave shapes tried. A range of sine waves from 400 to 4,000 cps was found to be suitable from both the engineering and the listener viewpoints. This range was later revised to 350 to 3,520 cps. The latter tuning combines equal logarithmic intervals with established musical pitches, inasmuch as the tones are separated by the musical interval of a major third. The slightly lower frequency range also was found to be more satisfactory for one older individual with hearing loss who was trained to use the device. Other subjects had no difficulty with the lower range.

The most important question remaining was the one of how many channels or tones should be used. A breadboard model of the projected reader was constructed, in which the number of channels could be varied from 1 to 12. Previous analyses of spectrographs constructed for each number of channels suggested that the highest number of tones should

FIG. 19-4 *Print-sensitive probe of the Battelle reader.*

yield the best letter discriminability. Somewhat surprisingly, no significant differences in letter identification could be detected when the number of channels was varied from 8 to 12, for either mechanical or simulated manual print tracking. Perhaps signal complexity in some way nullified the gains of increased letter resolution derived from a larger number of tones. Therefore, based primarily upon graphic studies and engineering considerations, 11 tones were provisionally adopted. The top and bottom tones were reserved for activation during tracking error, leaving 9 tones effective for the aural representation of printed letters.

One model of the Battelle reader weighs 9 lb and is portable. The print-sensitive probe or stylus (see Figure 19-4) is connected by a flexible cord to the chassis, and output sounds are presented by a headset. (A loudspeaker may also be used.) At the bottom of the stylus may be seen two parallel rollers that revolve as they are drawn across a page. Experiments showed significant gains in print-tracking accuracy with the addition of the rollers. Barely visible between the rollers are two medical-type lens-end light bulbs used to illuminate the printed page. The hexagonal control nut at the top of the stylus regulates a size-of-type adjustment. The range of type size that can be accommodated is 9 to 27 point, or approximately ⅛- to ⅜-in. type height.

The intensity of illumination from the stylus light bulbs is controlled by a control knob on the chassis panel. There is also a combined on-off and sound-volume control. Two headset jacks are provided for listening-in and instructional purposes. A multiple-connector plug connects the chassis and flexible stylus cord. Eleven sensitivity-adjustment potentiometers permit fine regulation of the threshold for triggering each of the tones.

Figure 19-5 illustrates the interior of the stylus. Reflected letter images pass across a linear array of 11 cadmium selenide photocells.

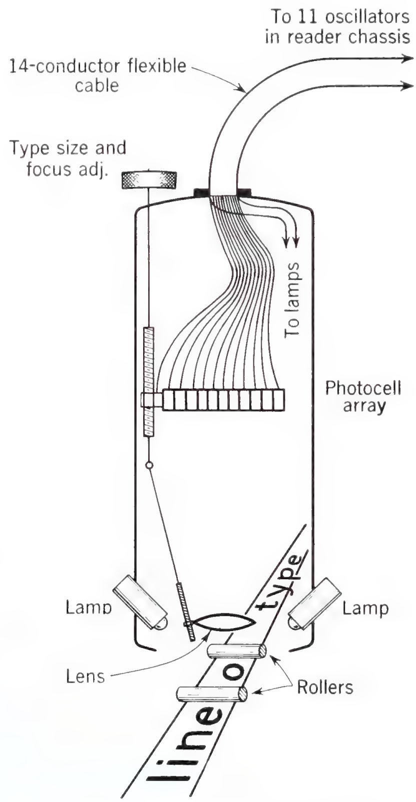
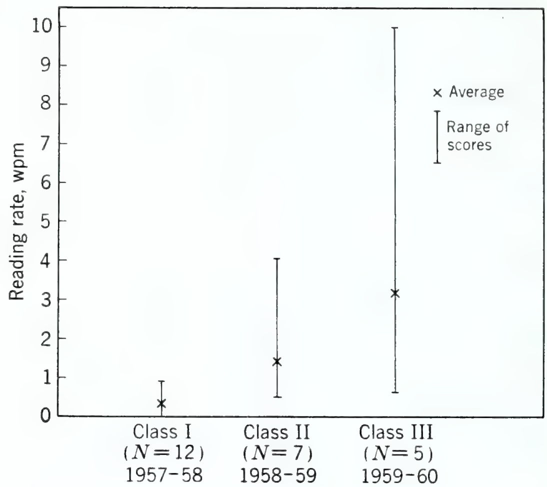
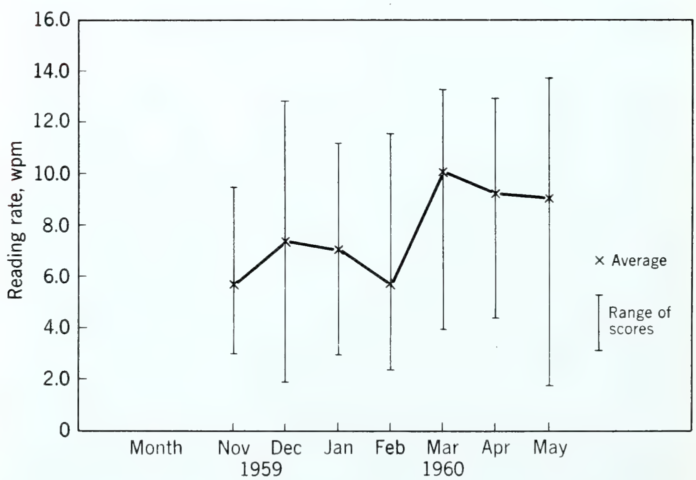


FIG. 19-5 *Schematic diagram of the reader stylus.*



(a)



(b)

FIG. 19-6 *Improvements in paragraph-reading rates with better machines, better training methods, and increased practice. (a) The beginning classes, each after approximately 100 hr of training; (b) four advanced students in 1959–1960 (three from Class I beginning in 1957–1958 and one with 1 year and a summer, 1958–1959). (From H. Freiberger and E. F. Murphy, Reading Machines for the Blind, IRE Trans. on Human Factors in Electronics, vol. HFE-2, no. 1, pp. 8–19, March, 1961. By permission of the publishers.)*

When a letter image covers one-third or more of any cell, the audio oscillator associated with that cell is turned on.

Blind subjects were brought into the development program as early as possible. In fact, the first class was trained largely from tape recordings before any reading instruments were completed. The effectiveness of this training method was reduced, however, by the extreme difficulty of adjusting to the reading instruments when these were introduced. Subsequent classes have benefited from an earlier introduction of the reading instruments and less emphasis upon tape recordings.

A recent article of Freiberger and Murphy [2] includes a summary of Battelle's training program. That summary is presented in part below:

"Several models of this instrument and of mechanical tracking devices have been developed. Major attention has been given to training and evaluation methods, with cooperation of blinded adults and the Ohio State School for the Blind. The subjects ranged in age from 15 to 60 years, with the majority around 18 years old. The gradual improvement in reading rates for paragraphs is shown in Figure 19-6*a* and *b* for successive classes of beginning students as well as over part of a year by a class of advanced students, the latter ranging from 2 to 14 wpm and averaging 10 wpm. Improved training methods, better devices, mechanical tracking, and selection for prescription should eventually raise the starting point, remove the lower ends of these ranges, and, hopefully, accelerate the trend toward 20 wpm.

"The inherent difficulties of the task are shown in Figure 19-7. Unfortunately, in most type styles as well as on the IBM typewriter used, many of the most common letters (e.g., e, c, s, o, a) are small, without distinctive ascenders or descenders, and similarly shaped in many respects. While ease of identification has improved since the early tests of these same subjects, there are still serious errors among some of the most frequently occurring letters. Technical improvements in the device and in a simple tracking aid are still needed to increase the consistency of signals.

"The confusion matrix, Figure 19-8, also offers clues for improved training. Even though 'e,' the most common letter in English, is the most difficult to identify in random-item tests, the errors consist chiefly of confusions with 's' (8 to 25 times) and with 'c' (4 to 25 times). Confusions typically are not reciprocal; for example, 'i' may be misread as 't' because most of the same channels are excited, but 't' (with all the upper channels excited, one near the top somewhat longer than the others) seems more distinctive. In general, context should help, because very few words would exist with such substitutions, and far fewer phrases or sentences would make sense. Here the human operator, if he can overcome under-

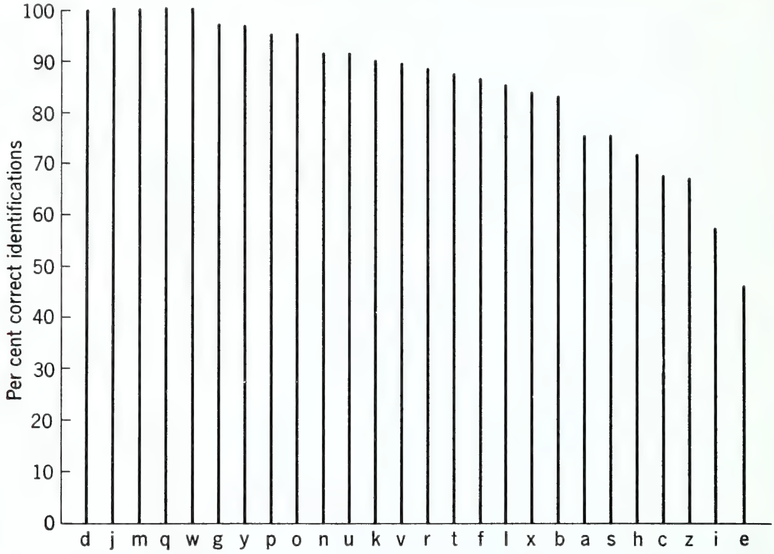


FIG. 19-7 *Accuracy of identification of lower-case letters by advanced subjects. (From H. Freiberger and E. F. Murphy, Reading Machines for the Blind, IRE Trans. on Human Factors in Electronics, vol. HFE-2, no. 1, pp. 8–19, March, 1961. By permission of the publishers.)*

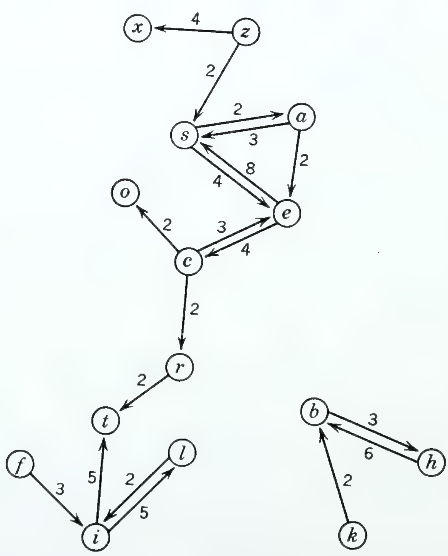


FIG. 19-8 *Confusion matrix for lower-case alphabet. (Each letter was presented twenty-five times. Only errors occurring two or more times were included in the matrix.) (From H. Freiberger and E. F. Murphy, Reading Machines for the Blind, IRE Trans. on Human Factors in Electronics, vol. HFE-2, no. 1, pp. 8–19, March, 1961. By permission of the publishers.)*

standable annoyance or even frustration, has a great advantage over a purely automatic reading machine in recognizing error, substituting the next most likely letter, judging success, and, if necessary, deciding to retrace. (Conversely, there is some risk, of course, that use of contextual cues may induce other errors by creating false expectations!)

"Emphasis is now being given to mechanical tracking methods and to training techniques stressing context. In addition, during the current year, the National Institutes of Health have begun to support additional studies of the possible usefulness of these devices for children. The Battelle device may prove adaptable to mounting on a typewriter to permit proofreading, erasing, and relocation of the carriage, or even immediate monitoring of touch typing. Such a system might have important vocational implications."

The reader also may be used to identify paper currency. A sequence of tests for the presence or absence of print at critical locations enables the user to identify bills.

Tape-recorded instruction can still be effective in certain situations. For example, a blind individual might be able to train himself at home with the aid of tape recordings. A series of tapes, giving verbal instructions as well as illustrative sound patterns, has been prepared for the self-training application. The tapes would be used with typed or printed practice sheets, and a reading device would be available for application of the recorded lessons.

At the outset of the development program, emphasis was placed upon individual-letter identification as the most direct test of the potential usefulness of the reading devices. However, with a given reading device, the question of training technique remains. Battelle is currently experimenting with both word- and letter-training groups. Designation as a word or letter group indicates an emphasis upon one or the other kind of training. A certain amount of both kinds of training is used to some extent for both groups. Preliminary findings indicate a slight superiority for the word group.

The approach used with the current letter group is comparable with previous training except for the tracking method used. Mechanical tracking aids are now used in the training, whereas previous classes relied almost entirely upon manual tracking. With manual tracking, a group of five students averaged 3 wpm after 100 hr of training (Figure 19-6*a*). With mechanical tracking, a group of five students now averages 7 wpm after 85 hr of training. This finding has led to the development of mechanical tracking accessories. The emphasis has been on simple devices, but a wide range of aids is being considered. The most complex device is pictured in Figure 19-9. This is a spring-loaded device with viscous damping. The spring is loaded by drawing the carriage to the left. Releasing

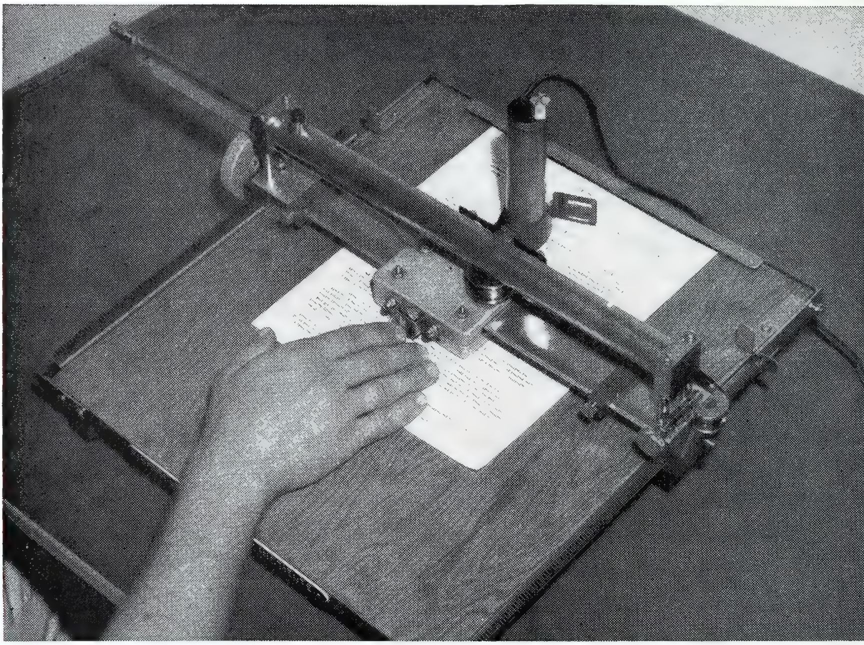


FIG. 19-9 *An experimental print-tracking device.*

the carriage results in a smooth and adjustable rate of return to the right edge of the page. Study is continuing to determine the best mechanical tracking device for general use. Even with mechanical tracking accessories provided, the manual tracking mode of operation will be retained as an alternative, depending upon the particular wishes of the user.

Results with the Battelle reader have been encouraging with some individuals, but less so with others. Some blind people who appeared interested and qualified in all respects failed to achieve any useful measure of skill with the device. Battelle has been unable to establish a correlation between success with the device and a great number of variables, including age, sex, date of onset of blindness, previous sighted-reading experience, measured intelligence, apparent motivation, or musical ability. Nevertheless, the search will continue for prescriptive criteria. Perhaps a test administered early in the training course would be a reliable indication of ultimate achievement.

Additional work is planned in almost every major aspect of the program—the development of improved devices, mechanical tracking aids, training methods, and the auditory code. A digital computer is being applied to both the analysis and the simulation of code variants. Codes are being investigated, with the goal of securing maximum letter differentiation and reading rates.

BIBLIOGRAPHY

1. d'Albe, F.: On a Type Reading Optophone, *Proc. Roy. Soc.* (*London*), vol. 90, ser. A, no. A619, 1914.

2. Freiberger, H., and E. F. Murphy: Reading Machines for the Blind, *IRE Trans. on Human Factors in Electronics,* vol. HFE-2, no. 1, pp. 8–19, March, 1961.

3. Heasly, Clyde C., Jr.: Report of the Subcommittee on Direct Translating Home Readers, Veterans Administration, New York, Aug. 3, 1956. [Also reported at the Fourth Technical Session on Reading Machines for the Blind, Washington, Aug. 23–24, 1956.]

4. Zworykin, V. K., and L. E. Flory: An Electronic Reading Aid for the Blind, *Proc. Am. Phil. Soc.*, vol. 91, no. 2, pp. 139–142, 1947.

Chapter 20

TOWARD A HIGH-PERFORMANCE READING MACHINE FOR THE BLIND

Franklin S. Cooper[*]

An important factor to be considered in the design of a reading machine for the blind is the intended use of the device and the level of performance needed to realize that use. For rapid and extensive reading, the performance of the machine must approach that of a sighted reader speaking at a normal rate. That speech itself is a requirement for high performance may appear surprising; however, there is a solid basis in theory and experience for such a conclusion [1]. This is not to say that a high-performance reading machine must speak perfect English, but rather that it must produce understandable English at not less than normal rates.

There are at least two reading-machine systems now under development that are aimed at this goal.[1] One generates a spoken output by rearranging voice recordings of individual words into the sequences found in printed text; the other generates synthetic speech on the basis of letter-by-letter information from the text. In both systems, the first step is recognition of the printed characters—no mean task when various type faces and page formats must be accommodated. In one machine the compilation of a voice output from stored recordings of spoken words requires a large random-access memory; in the other, synthesis of the output speech requires sophisticated instrumentation and procedures for coping with the vagaries of English spelling.

There is no immediate prospect of an inexpensive, portable device adapted to personal use. A high-performance machine will probably be fairly large and expensive, designed for institutional use in the preparation of tape recordings or for use by individuals only in the context of a reference library. The inherent complexity of high-performance reading

* Haskins Laboratories, New York, N.Y.

[1] Reading-machine research is being conducted at the Haskins Laboratories for the Research and Development Division, Prosthetic and Sensory Aids Service, Veterans Administration, under Contracts V1005M-1254, Output Characteristics and Construction of an Interim Word Reading Machine, and V1005M-1253, Research on Audible Outputs of Reading Machines for the Blind.

machines requires that research and development proceed slowly and that one must often speak of plans, problems, and reasonable expectations. Nevertheless, there is progress to report and the remainder of this chapter concentrates on the design of an interim word-reading machine. The chapter also deals briefly with the requirements for a full-scale machine of this kind. A basis for the design of machines that synthesize speech from letter information is discussed elsewhere [3].

A WORD-READING MACHINE: GENERAL CONSIDERATIONS

The objectives to be met in the design of a full-scale word-reading machine are implicit in its name: it should scan a wide variety of printed or typewritten material and emit a sequence of spoken recordings of the words of the text, preferably in a lifelike way. A partial solution of the reading problem may soon be attainable, since suitable subassemblies are under intensive development for business applications of electronic data processing. Such a machine must, however, be taught how to talk. Indeed, the level of its performance and the level of satisfaction to be experienced by the user will depend almost entirely on finding a successful solution to just this "language problem," i.e., how to provide recordings of individual words that can be shuffled about like movable type and yet flow into the accustomed patterns of connected speech. This is the problem with which this discussion is principally concerned. However, let us first consider briefly the hardware requirements.

Scanning and character recognition are obvious and necessary functions of the reading machine; they differ from the functions involved in handling business documents only to the degree that the reading machine must handle a wider range of formats and type fonts. For scanning and character recognition, first-generation business machines use special means such as magnetic ink and special character shapes; however, one may expect a later generation of less specialized devices that will meet the blind user's need for access to a wide range of printed and typewritten material.

The other major hardware component is a high-capacity random-access memory. Here again equipment designed for electronic data processing should suffice, although the detailed requirements of a reading machine differ somewhat from those usually encountered. In a reading device the limitation on access time is not very severe, since a reading rate of 180 wpm, a fairly lively pace, allows on the average $\frac{1}{3}$ sec per word. Some words, to be sure, require a much shorter time than this and cause additional limitations on the dead time that can be permitted be-

tween words that normally cluster together to form a phrase. With some short-term buffering of the output, however, the access requirements could readily be met by disk or drum memories, though not by magnetic tape for real-time operation. The requirements on memory capacity are harder to estimate, since they involve interdependent design considerations. As a result, it will be more appropriate to consider the question of memory size after a discussion of the choices made in the development of the particular prototype machine under study.

AN INTERIM WORD-READING MACHINE

The need for an interim device arises directly from the language problem, that is, the one area not covered by existing data-processing developments. Actually there are many and various language problems. Some, such as the extent and organization of the vocabulary, might seem to involve only a judicious compromise with memory size. However, the price of this compromise is interruptions in the reading to allow the machine to spell words that are not in memory. It is possible to estimate the frequency of these interruptions, but not the annoyance that they can cause. Prolonged listening to texts that have controlled proportions of spelled words may be the only feasible way to assess this factor and so secure a firm basis for decisions on memory size.

The need for exploratory research is even more evident when one considers such general problems as intonation, tempo, and phrasing or such specific problems as choosing the pronunciation for particular words and the sounds, or lack of them, to replace punctuation marks. Investigations of sample texts can be helpful, but the reaction of actual blind users to extensive test materials will be a surer guide.

It is evident that research on the language problem is a necessary preliminary to the design of a word-reading machine of any kind. This research should include the production, for field tests with blind users, of substantial amounts of output text in which the language variables are manipulated.

The present research program on word-reading machines began with the above premises and has three overlapping phases. The first consists of experiments for testing various design possibilities. The second phase of the program is the design and construction of a machine adequate for preparing texts but otherwise as simple as possible.[2] Finally, the interim

[2] A number of paragraph-length texts have been prepared by hand methods using a modified Languagemaster to read single-word recordings on cards and to re-record these words on magnetic tape with a start-stop recorder. The trial recordings show (1) that it is feasible to generate connected speech at nearly normal

word-reading machine is to be used to provide a weekly news service for a group of blind persons so that their reactions to the intentional manipulation of certain language variables can be followed through regular interviews.

The device that is now nearing completion is intended to test many aspects of the language problem and to be equivalent to a full-scale word-reading machine in the quality of its speech output. It is an interim device only in the sense that it bypasses such major engineering problems as character recognition and real-time access to a large memory. Thus, Teletypesetter (TTS) tapes, available as a by-product of the publishing industry, provide input material that would otherwise come from character-recognition equipment operating on printed text. The need for fast access to a large memory is likewise evaded; the voice recordings of successive words are found, one by one, by searching through a multi-track magnetic tape that has paired entries of a digital address and a voice recording for each word of the vocabulary.

The recorded words are not read out immediately, as they might be from a real-time memory, but are transferred to a start-stop recorder that can wait as long as it must for the next word to be found. Thus, the interim word-reading machine can prepare, over a period of hours, a spoken version of the text that was originally punched into the TTS tapes and thus permit a thorough evaluation of the roles of various language variables. Finally, after its research assignment is completed, the interim machine can be used to provide blind listeners with a weekly news service recorded from by-product TTS tapes provided by one of the major news magazines.

The operation of the interim word-reading machine is illustrated in Figure 20-1. The input tape is read into temporary storage in the decoding unit. Since the TTS code includes not only the desired text but also instructions for type casting, each character must be interpreted before it is used or before it is rejected and another character is called in from the tape reader. Consecutive letter codes are put into a 60-position shift register in the scan-and-compare unit. Tape reading is stopped when the end of a word is reached, as signaled by *space, carriage return,* or the indication of a punctuation mark. There are, to be sure, some complications. For example, *hyphen* usually signals a compound word to be handled either as one unit or as two successive words, depending on which procedure proves to be more advantageous; however, *hyphen*

---

reading rates and yet retain high intelligibility and a measure of naturalness but (2) that maintaining this level of performance requires careful control of the production and recording of the separate words. The discouraging results reported by A. N. Stowe and D. B. Hampton [5] are interpreted as an indication of the difficulty of adequate control and the complexity of the language problem.

followed by *carriage return* means that the preceding letters constitute only part of a complete word. Contingent interpretations of this kind of signal call for temporary storage of several successive characters and a certain amount of logic between the decoding tree and the shift registers.

The dictionary search, initiated by an end-of-word indication from the decoding unit, can now proceed. The shift register of the scan-and-compare unit contains the address of the desired word in the form of the digital code for its letters, or the first 10 characters if the word is a long one; the identity of the first letter of the word has been used by the system-control unit to select one of the 14 available pairs of tracks on the dictionary tape. One track of each pair contains the digital addresses of

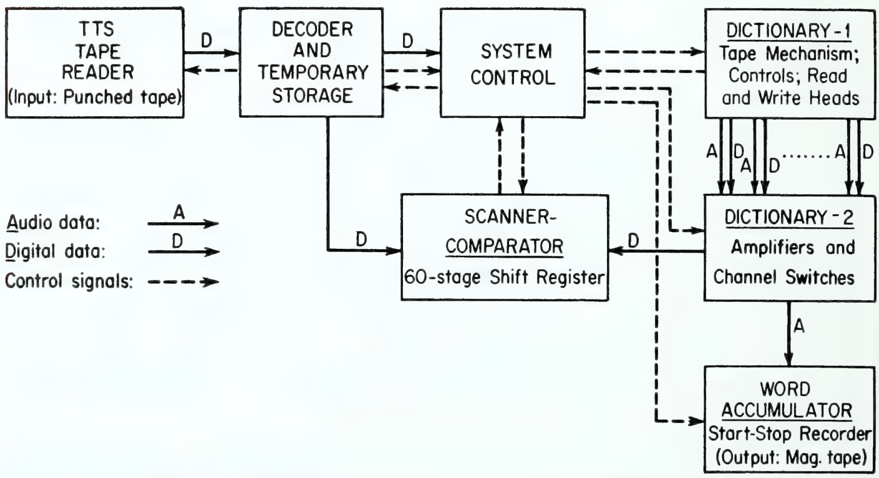


FIG. 20-1 *Block diagram of an interim word-reading machine.*

words that begin with the same letter as the target word stored in the shift register; the other track contains the corresponding voice recordings. The search proceeds at high speed, with the digital addresses from the dictionary tape being compared, bit by bit, with the target address stored in the scan-and-compare unit. The shift register of this unit is advanced and synchronization is checked, or reestablished if necessary, by clock and framing pulses from two additional tracks on the dictionary tape.

An exact match between the two addresses means that the desired word has been found. Accordingly, the tape transport of the dictionary unit is shifted abruptly from fast forward to slow reverse in order to transcribe the voice recording onto the ¼-in. tape of the word-accumulation unit. This transport is started by a tone burst in the voice track of the dictionary tape and stopped by the next framing pulse, thus allowing

an appropriate duration for the audio version of each word even though a fixed frame length is used to simplify the search procedure. It now remains only to return the dictionary tape to its home position, to clear the shift register, and to initiate the reading of the next word from the TTS tape. The entire cycle requires, on the average, about 10 sec to yield about ⅓ sec of speech; i.e., the device operates at about one-thirtieth of real time.

What happens if an exact match is not found? Since the words in each track of the dictionary tape are ordered by word length, the search for a word that is n letters long need proceed no further than the first word of length $n + 1$. There is the additional advantage that search time is reduced still further because the words that are used most frequently, usually the shorter words, are examined first. Failure to find the target word means, in the simplest case, that it must be spelled by using the successive letters stored in the shift register.[3] An intermediate step is planned, one that would certainly be needed in a full-scale machine: failure to match the address in the main dictionary tape would initiate a second search in a track pair reserved for specialized vocabularies. Thus, if it were missing from a general dictionary of medium size, "violin" would be found in a specialized vocabulary containing musical terms. The interim machine now under construction cannot take full advantage of this procedure unless another transport mechanism is added for specialized dictionary tapes; it does, however, permit an adequate test of the usefulness of specialized vocabularies as supplements to the main dictionary.

The instrumentation of the interim word-reading machine is fairly conventional. A Friden paper-tape reader transmits the TTS characters directly to a relay decoding tree and a transistorized shift register for temporary storage. The main shift register, the various logic units for comparison and control functions, and the record and playback amplifiers employ transistorized circuits on plug-in cards. The tape-transport mechanisms for both the main dictionary and the word accumulator are fast start-stop units that move their tapes from bin to bin. The inch-wide dictionary tape is searched at 60 in./sec or read at 3¾ in./sec. The search can proceed in either direction from a home position at the middle of the tape, thus doubling the number of track pairs that can be used. Separate drive assemblies are used for the two speeds, and an electrical interlock prevents operation of more than one of the four pinch-roller solenoids at the same time. Separate stacks of eight heads each are used for interlaced audio and pulse signals. The heads have individual preamplifiers that precede the switching relays used to select a particular

[3] The form of spelling might be that of spelled speech, described by Metfessel in Chapter 21 of this book.

track pair. A total tape length of 250 to 300 ft can accommodate the design objective of a 6,000-word vocabulary.[4]

In short, the whole design was aimed at minimum engineering complexity and development cost, together with realistic simulation of the operation of a full-scale word-reading machine.

MEMORY REQUIREMENTS FOR A FULL-SCALE MACHINE

Some of the factors that will determine the memory size of a full-scale word-reading machine have already been noted. (1) The interruptions for spelling will be less frequent and less irksome if a large memory is used, but access to such a memory may prove slow or expensive. (2) It may be possible to gain many of the advantages of a large memory by using appropriate specialized vocabularies to supplement a main vocabulary of modest size. (3) The requirement on speed of access for real-time operation can probably be eased substantially by the use of buffer storage. It may be possible also to obtain more efficient use of a given memory by resorting to certain artifices. Thus, dual entries for the many words whose plurals are formed in "s" or "es" can probably be avoided by adding a fricativelike sound after the voice recording of the singular form—a departure from naturalness, but probably not a very objectionable one. Some other suffixes might be processed similarly, though at greater cost in machine logic and a smaller saving in memory size.

A major consideration in choosing a memory is the form in which the recordings are to be stored. The addresses of the words pose no problem, since digital representation is both logical and economical. Storage of the voice recordings is quite another matter. Analog recording of the acoustic waveform was an obvious choice for the magnetic tapes of the interim word-reading machine. Conventional techniques permitted about 30 words per linear foot of 1-in. tape, even though 9 of the 16

[4] A vocabulary of 6,000 words was chosen as a design compromise among several factors: complexity of tape-handling equipment, cost of recording the dictionary tape, and adequacy of the vocabulary as indicated by the frequency with which missing words would have to be spelled. Some idea of the trading relation between vocabulary size and percentage of running text that would remain to be spelled can be had from these rather rough estimates: 50 per cent spelling rate for a vocabulary of 100 words; 25 per cent for 1,000; 10 per cent for 3,000; 5 per cent for 6,000; 1 per cent for 15,000 to 20,000. The number of different words in Webster's Collegiate Dictionary is about 60,000; more than 600,000 are claimed for Webster's New International Dictionary (2d ed.). The spelling rate can be reduced by using specialized vocabularies to supplement a main vocabulary, but the effect is hard to estimate. It seems unlikely, however, that fewer than 10,000 to 20,000 words will be required.

tracks were used for address and control functions. Despite the advantages of combining analog and digital techniques, it may well happen that the best available memory for the full-scale machine will be a strictly digital one. Of course, the speech could be put into binary form, through the use of pulse code or delta modulation, but the cost in bits is high. For example, an average five-letter word that lasts ⅓ sec requires only 25 bits for its address but 10,000 to 15,000 bits for a voice recording of telephone quality. At this rate, a 10,000-word vocabulary would require a memory capacity of about ¼ million bits for the addresses and well over 100 million bits for the audio. Such a requirement could be met, but only by the high-capacity disk and card systems now being announced as computer accessories or by photographic or other types of memories still in the experimental stage.

Can acceptable speech be had for fewer bits? Since the quality of pulse-code-modulation speech deteriorates rapidly with decreasing bit rate, some other way must be sought to store and regenerate a digital speech signal. The obvious method is speech synthesis, using the digital information to control a synthesizer rather than to reconstruct the waveform. This is, in fact, a rather attractive solution, for speech of adequate intelligibility can be obtained at rates of about 2,000 bits/sec by analysis-synthesis methods. Speech quality is marginal, but the task of recording the memory would be no greater than that by analog methods. Gunnar Fant and his co-workers have demonstrated synthetic speech that is almost indistinguishable from the spoken version despite bit rates as low as 1,100 bits/sec [2]. Control of the synthesis was based on painstaking measurements of real utterances. The construction of a large vocabulary by these methods would be a formidable task indeed, but semiautomatic methods may be developed, and a controlled synthesis would have the advantage of allowing the precise manipulation of intonation, intensity, and so forth. The memory capacities required by this method would be about 400 bits per average word or 4 million bits for a 10,000-word vocabulary,[5] which is well within the capabilities of the larger drum and disk memories now in use.

There is an obvious resemblance between word-reading machines of this kind and the machine that synthesizes its output speech on a letter-by-letter basis. The essential difference lies not so much in the hardware as in the choice of linguistic unit for controlling the synthesis.[6] Certain consequences follow from this choice: the word machine will

[5] This begins to approach, perhaps to within a factor of 2, the storage-area requirements for analog representation of the speech waveform without requiring a memory device that can handle both analog and digital data.

[6] For a discussion of the factors affecting this choice, see References 3 and 4 and the references to earlier works cited in these two articles.

produce the more lifelike speech (except when it must spell); the letter machine will be intelligible, and it will never need to spell, but its dialect may often be as bizarre as the English spelling on which it depends. In hardware terms, the word machine requires a large memory, whereas the letter type needs little memory but much logic; of the two machines, the letter type will probably be the less complex and the less costly. Fortunately, a large part of the research that is done on either system will be applicable to the other and will help to clarify the complex factors on which a choice between the systems must be based.

Since high performance implies that a reading machine for the blind must be able to "speak" intelligible English (or other natural languages), one must expect the device to be complex and its development laborious. However, technological advances in electronic data processing, combined with research under way on the language problems of teaching such devices to talk, offer promise for a useful library machine within a few years. Whether the machine will speak from recordings or by synthesis is not yet evident, but that it will speak seems certain.

BIBLIOGRAPHY

1. Cooper, F. S.: Research on Reading Machines for the Blind, in P. A. Zahl (ed.), "Blindness: Modern Approaches to the Unseen Environment," pp. 512–543, Princeton University Press, Princeton, N.J., 1950.

2. Fant, C. G. M., et al.: Recent Progress in Formant Synthesis of Connected Speech, paper presented at Sixty-first Meeting of the Acoustical Society of America, Philadelphia, May 10, 1961.

3. Liberman, A. M., et al.: Minimal Rules for Synthesizing Speech, *J. Acoust. Soc. Am.*, vol. 31, pp. 1490–1499, 1959.

4. Peterson, G. E., and Eva Sivertsen: Objectives and Techniques of Speech Synthesis, *Language and Speech*, vol. 3, pt. 2, pp. 84–95, 1960.

5. Stowe, A. N., and D. B. Hampton: Speech Synthesis with Pre-recorded Words and Syllables, *J. Acoust. Soc. Am.*, vol. 33, pp. 810–811, 1961.

Chapter 21

SPELLED SPEECH

*Milton Metfessel**

TWO VARIETIES OF VOCALIZED SPELLING MAY BE DIFFERENTIATED: DISCRETE spelling and spelled speech. With the former, the kind that ordinarily occurs, each letter is heard as a unit and identified separately. It is illustrated well by the procedure of a spelling bee. In contrast, spelling may be done so that letter sounds are run together and not identified separately. Sounding more like speech than spelling, it has been named "spelled speech." For each word there is a new pronunciation, which resembles a foreign-language word rather than a sequence of letters. For example, the spelled-speech version of ME is *Emmy* (ɛmi)[1] and of BUT, *Beeyouty* (biɪuti). These new pronunciations are formed from letter names, but they flow together so that the combination is perceived as a configuration rather than as a number of individual elements.

For automatic readers that would make use of spelled speech in the output, there would be stored in the speaker unit 26 letter sounds. These would be played one after the other in various combinations to spell words as the scanner moved across a line of print. Spelling would thus be produced synthetically, all words being formed from a single set of 26 voice fragments.[2]

Synthetic production of discrete spelling is not difficult, since the various letters do not have to coalesce. To synthesize spelled speech, however, it is necessary to find a set of alphabet sounds each of which will go smoothly with every other letter that it may adjoin in the English language. To select such letter sounds has been one phase of the spelled-speech research.

* Communications Laboratory, Psychology Department, University of Southern California, Los Angeles, Calif.

This research was carried out under Contract V1005M-458 with the Veterans Administration, Prosthetic and Sensory Aids Service.

[1] Based on the alphabet of the International Phonetic Association.

[2] Natural spelling (i.e., any instance in which a person actually spells material) is not limited to 26 voice fragments. Each letter is produced anew whenever it occurs in the text. The speaker may vary pitch, intensity, timbre, and/or duration of any letter from one time to another so that it will go smoothly with adjoining sounds.

DEVELOPMENT OF ALPHABETS FOR SYNTHETIC SPELLED SPEECH

The goal has been to develop one or more sets of recorded alphabet sounds that, if used in an automatic reader, would provide:

1. Easy identifiability
2. Rapid rate (90 to 180 wpm)
3. Coalescence (that is, smooth transitions from letter to letter so that words and phrases rather than separate letters become the units of perception)
4. Acceptable sound quality

The general procedure has been (1) to have individuals practice spelling in which letters are run together, (2) to record their productions on tape, and (3) to select letter sounds on the basis of the above criteria. During the research, more than 20 alphabets have been constructed, each closer to the goal.

The accurate and almost immediate perception of slowly produced letters was no problem. The question was the extent to which letter length could be decreased without loss of intelligibility. Under ordinary circumstances it takes much longer to spell a word than to pronounce it because there is now one syllable for every letter. The usual rate of spelling (10 to 25 wpm) was considered undesirable for an automatic reader because possible users would be less likely to be motivated to acquire skill if maximum reading rate were so limited. With spelled speech, in contrast to discrete spelling, higher rates are possible because letters follow one another more closely and because there is more rapid production of letters.

Two methods of reducing the length of letter sounds were used, sound contouring and natural compression. The former involved (1) removing for each letter the central section of tape that involved repetitiveness of sound waves and (2) splicing together the end sections, characterized by less redundancy. Location of segments to be removed was determined from both sound (when tape was played slowly) and visual representation of waves (as revealed by a method developed in the laboratory).[3] Natural compression consisted in having individuals

[3] If a section of magnetic tape on which sounds have been recorded is rotated through a mechanically activated solution of iron powder and a highly volatile fluid such as carbon tetrachloride or freon, the iron particles will collect at points of magnetization. The result is a visualization of the sound patterns on the magnetic tape. If transparent cellophane tape is pressed firmly upon the magnetic tape, the iron particles are mechanically transferred to the cellophane. When it is pulled away from the magnetic tape, the cellophane tape has both visual and auditory properties. The sounds can be reproduced either by a magnetic type of circuit or by an optical

practice spelled speech with emphasis on speed. It was demonstrated early that, with these two methods, alphabets meeting the criteria of easy identifiability and a spelled-speech rate of 90 to 100 wpm could be constructed with relative ease. Subsequent work concentrated on letters that would provide this speed. Identifiable letters that would permit spelled speech up to a rate of 180 wpm were also obtained in this early work, but they were of less satisfactory quality.

The requirement for good sound quality, important in connection with motivation of the user, covers not only the sound of the voice itself but also avoidance of a monotonous effect in lengthy sequences of spelled speech. Concern with the former led to making test recordings of numerous voices and selecting only those recordings whose voices were judged pleasing. The most obvious way of avoiding monotony, by introducing among letters considerable variation in pitch and intensity, was found to be associated with poor coalescence. It has, however, been possible to identify subtle variations that do not produce rough transitions. A slight drop in pitch and/or intensity at the end of letters most likely to occupy the final position in words is one example. Furthermore, introduction of rhythm or syncopation by differential spacing of letters has been found to prevent monotony in spelled speech.

Coalescence

When synthetic spelled speech was constructed with early alphabets, markedly varying results were obtained. Some combinations sounded like natural spelled speech. With others, echolike effects were heard. Many combinations sounded as though they were spoken by two persons.

Some factors associated with poor coalescence were subsequently found to be due to the method of combining letter sounds, i.e., by splicing of tape, which would not be present in an automatic reader. This led to exploration of other methods for producing synthetic spelled speech so that it could be studied prior to construction of a reader. One system involved putting each letter on a loop of tape so that it could be recorded over and over at regular intervals. A second letter was then recorded on the same tape from a loop differing in length. The disparity in loop lengths produced systematic variation in amount of space between the two letters, and when spacing was such that one letter followed immediately upon the other, the relative coalescence

---

system (used in motion pictures) with a light passing through the moving tape to a photoelectric cell. A spray coating of clear plastic is used on the cellophane tape and may also be applied to fix the iron particles on the magnetic tape if use is not to be made of the cellophane-tape procedure.

of the combination could be determined. Subsequently, more effective splicing procedures were also devised.

However, smooth transitions require more than an effective method of putting letters in juxtaposition. In trials with various alphabets, much detailed information was obtained concerning letter characteristics important for coalescence. For example, those letters with carrier vowels varying markedly in pitch or intensity were, in general, found to coalesce poorly. It was also discovered that letters with certain phonetic characteristics behaved in similar fashion. For example, consonants with final carrier vowels of *ee* (i), as in the letter T, following other letters were found to coalesce poorly if the start of the consonant was not separated by a small time interval (approximately $\frac{1}{60}$ sec) from the end of the preceding letter. A similar time interval was found to interfere with coalescence in the case of other letters. Information of this kind was formulated in terms of objective criteria to be applied in choosing letters. Particular attention was given to developing means of predicting whether a given letter would coalesce with others.

The general system devised for selecting letters involves using phonetic equivalences (1) to reduce the number of letter samples that need to be considered and (2) to predict what will happen when various letters are placed in combination. In applying this system, only those letter samples are considered which are short enough for the desired rate and which can be distinguished from other letters.

1. The first step is to select a T, final carrier vowel *ee* (i), and an S, initial carrier vowel *eh* (ε), that go together well when combined to form either ST or TS. If the recording session includes words with ST or TS combinations, half this step will be taken care of by the natural coalescence of the two letters in one order or the other. The letters selected are also tested to be sure that they coalesce well with themselves, that is, that the SS and TT combinations are smooth ones. This double-letter test is applied to all letters as they are considered, for it has been found that, if a letter does not provide a good double-letter combination, it has small chance of going well with other letters.

2. The second step is to select samples of the vowels and combine them with each of these two letters, getting the combinations AS, SA, AT, TA, ES, SE, ET, TE, etc. If it is not possible to find a sample of each vowel that will go smoothly when placed before and after the S and T that have been chosen, other S's and T's that coalesce need to be substituted for the ones originally chosen until a satisfactory result is obtained. If A coalesces with T and S and if E coalesces with T and S, then the prediction may be made that A will coalesce with E. Similar predictions may be made for the other vowels.

3. The third step is to select samples of all other letters with carrier

vowels of *ee* (i) that will coalesce with the S that has been selected. In similar fashion, selection is made of all other letters with initial *eh* (ε) sounds on the basis of their coalescing with the T that has been chosen. Since the S and T coalesce and since all letters phonetically equivalent to S have been matched to the T and those phonetically equivalent to the T have been matched to the S, the prediction may be made that any one of these letters will coalesce well when combined with any other.

4. A similar procedure is then applied with the remaining letters of the alphabet.

This method was tested by using it in constructing the most recent alphabet. A highly skilled speller was asked to spell a specially selected list of words a number of times in a single session. The list consisted of 26 words that, in the speller's previous recordings, had provided the best sound for each letter. Including all in a single session, it was predicted, would make possible a new alphabet with improved coalescence. This prediction was confirmed, and the current alphabet was judged to meet satisfactorily all criteria.

Work has begun, also, on a new concept of spelled speech. Class II spelled speech (paralleling Class II braille) is an abbreviated form that would increase the speed possible with a reader. Its basic idea is that the output of the machine would differ for some letters in terms of the letter by which they were followed. For example, P, T, B, and D when followed by a vowel or by any letter with an initial carrier vowel (such as F, H, R, X) would have most of their carrier E excluded. Thus, TH would be sounded not *Teeaich* (tieɪtʃ) but *Taych* (teɪtʃ).

TRAINING IN SPELLED SPEECH

The Training Series

Since spelled speech is not a customary method of communication, a major question is that of how to train individuals to receive such material with minimum time and effort. The task may be described as learning to read with one's ears. To work out effective training procedures, three major training series were conducted with four groups of college students, each with close to 20 persons:

1. 19 sessions, about ½ hr in length, 2 per week
2. (*a*) 13 sessions, about ½ hr in length, 2 per week; (*b*) 7 sessions, about 1 hr in length, 1 a week
3. 5 sessions, at intervals of 1 week or more, each lasting about ½ hr (taped material 15 min)

Sighted rather than blind persons were used in this preliminary work, with the presupposition that training methods suitable for the former would be appropriate for the latter. The skill developed by some of the subjects in understanding spelled speech is the more remarkable in view of the fact that the largest amount of formal training that any subject received was limited to about 10 hr.

Some of the material presented consisted of alphabets being developed and synthetic spelled speech produced with these alphabets. However, it was felt that restricting sessions to synthetic material available at the time would prevent exploring fully the extent to which spelled speech could be mastered. Consequently, natural spelled speech was also utilized. This procedure involved the presupposition that there would be no basic differences in effective training procedures for (1) synthetic spelled speech constructed with a satisfactory alphabet and (2) natural spelled speech. Almost all material was presented by playing it on a tape recorder.

The material included single letters, bigrams, trigrams, words of varying lengths, phrases, sentences, questions, and a 10-min skit. With most of these, the subjects' task was to identify what they heard. However, they were also instructed to answer the questions, and, following the skit, they were queried concerning its content in conventional English. This approach was used to explore how well they could grasp the meaning of material given in spelled speech. Numerous other variations in procedure were also investigated.

Findings

The following recommendations regarding effective training methods were made on the basis of the results of these training series:

1. Individual rather than group training.

2. Participant control of presentation of material, so that whatever the subject does not receive on first exposure can be repeated as often as necessary for learning it. (The subject initiates presentation of the letter by throwing a switch.)

3. Recording by means of taping verbal responses, in order to eliminate restrictions on progress that are associated with having to write responses. Such a method permits utilization of speed of response, in addition to accuracy of response, as a criterion of learning.

4. Presentation of material in such fashion as to give immediate knowledge of results. (The subject hears the correct response 2 sec after he gives his own response.)

5. Restricting alternatives for the subject by means of the sequences of items spelled and the instructions used.

In addition to laying the foundation for standardization of training procedures, results from the series provided evidence in support of the following:

1. Learning to understand an output composed of synthetic spelled speech is probably much less difficult than learning arbitrary tone patterns, which have been tried as output for other types of readers. In learning ordinary spelling, individuals have built a foundation that would be expected to have a positive transfer effect in the learning of synthetic alphabets. That this is the case is indicated by the finding that, even with synthetic alphabets regarded as far from the best achievable, there was high accuracy of identification on the first trial. For example, in their first session the subjects of group 1 listened to a recording of letters of one of the synthetic alphabets, and after hearing each letter they recorded what they thought it was. Since one letter was presented twice, the maximum score possible was 27. The median number of correct responses on the first trial was 25; the lowest score made by any person during the session was 20. This high level of performance is to be contrasted with the weeks of training time required to familiarize subjects with a new alphabet of arbitrary tone patterns.

TABLE 21-1 *Communication of Meaning in Session 4 for Training Group 3*

Item presented in spelled speech	Number of subjects giving meaningful response
THE TIME	11
YOUR FIRST NAME	6
TODAY'S DATE	10
YEAR OF YOUR BIRTH	5
WHERE YOU ARE NOW	9
YOUR MAJOR	9
FAVORITE FOOD	16
KIND OF CAR YOU PREFER	6
COLOR OF YOUR EYES	8
YEAR YOU GRADUATE	5

Increase in the accuracy of response to words with repetition was found in the training series. For instance, the material in the second session for group 3 involved 110 words, 14 of which were presented four times. The median percentage of correct identifications for these was 70 on the first presentation, 86 on the second, 90 on the third, and 94 on the fourth.

2. With relatively short periods of training, individuals are able to respond to the meaning of material given in spelled speech. Results from the fourth session for group 3 are cited in illustration. Seventeen subjects were asked to make a meaningful response to 10 items given in natural spelled speech. Table 21-1 lists these items and shows the number

of participants able to perform the task. The students' scores ranged from 0 to a perfect score of 10, with a median of 5. It is to be emphasized that this set of results was obtained after the participants had had only three training periods, a total of about 45 min. On the second presentation, more than half the subjects gave a meaningful response to each item, and the number of correct responses increased for every item except one (FAVORITE FOOD).

The rate of learning shown in these training sessions and the high level of achievement attained by some of the participants after only a few hours of training leave little room for doubt that spelled speech can be mastered sufficiently to make it practical as output for a reader. As in most learning situations, not all persons performed equally well. Whether performance differences stemmed from variations in motivation or ability or from the suitability of training methods is not known. Questions regarding this problem and the maximum reading speeds to be anticipated with lengthy spelled-speech training can be answered only by additional research.

Chapter 22

ELECTRONIC TRAVEL AIDS FOR THE BLIND

*T. A. Benham**

DESPITE THE EXCELLENCE OF BOTH THE LONG CANE AND THE DOG AS TRAVEL aids for the blind, relatively small fractions of the totally blind population use either. Although these proportions are likely to increase with the current development of rehabilitation centers specializing in mobility training, there appears to be some demand for a device that would not cause physical contact with other pedestrians, would be relatively inconspicuous (although some blind persons prefer an indicator of blindness such as a white cane or a dog as a warning to motorists), and would permit operation in crowded and noisy environments in which auditory cues are absent or useless.

Any mechanical method for detecting hazards during travel by the blind should fulfill at least the following three functions:

1. It should detect obstacles and roughly indicate their proximity. Large obstacles such as walls are easily detectable by auditory cuing, but obstacles such as chairs, thin posts, bicycles, and doors standing ajar cause considerable discomfiture to the unaided blind traveler. A special difficulty posed in traveling over familiar ground is the appearance of unexpected obstacles such as movable trash cans or pedestrians.

2. It should detect down steps and holes. Some holes, such as an open manhole, have some form of guard around them that is detectable by ordinary obstacle detectors. However, a down curb, a flight of stairs, or a sudden increase in the steepness of a hill poses special problems to the blind.

* Department of Physics, Haverford College, Haverford, Pa.

The research reported in this chapter was supported by the Veterans Administration under Contract V1001M-1900.

This is a summary of a longer, much more detailed report by the author, Electronic Obstacle and Curb Detectors for the Blind, Summary Report, Veterans Administration Contract V1001M-1900, June 30, 1960, which describes completely the experiments that were performed prior to the selection of and during the development of the obstacle and curb detectors discussed in this chapter. Copies of the major report may be obtained from the author.

3. It should detect up steps and low obstacles that may well be too small to be detected by methods used primarily for detecting walls and doors. In addition, some blind individuals desire a travel aid to help them walk in a straight line and to give compass bearings. However, there has been little serious study of this type of guidance device.

About 20 guidance devices, working on a variety of principles, have been built in the last 15 years. The majority of them, with details of operation and evaluation, are described in "Blindness" [10]. A few operate on ambient light reflected by the obstacle, but most radiate a beam of sound or electromagnetic radiation and detect the reflected ray. Most of the devices elicit information regarding range and the presence or absence of an obstacle. Most investigators have considered the detection of substantial obstacles such as walls or doors to be the principal goal of guidance devices. Gradually, however, they have realized the importance of step-down detection, which is more difficult than the detection of large, conventional obstacles, and of step-up detection, which presents further difficulties. Throughout the development of these devices, the greatest emphasis has been upon systems by which an able-bodied individual lacking only vision could find his way about safely at a walking speed of 3 or 4 mph. The individual's reaction time, stumble recovery, and general physical condition have been assumed to be normal or perhaps even somewhat highly developed.

Reflection of sound is the most versatile source of travel information available to the blind. Sound, as from the tap of a cane or the scuff of a foot, reflected from an object would announce the obstacle's presence to the acutely perceptive, but the newly blind or even the experienced traveler in a noisy street would find these echoes too faint to be of help. To overcome this problem, several workers have developed small, easily carried horns that generate high-frequency sound in sharp bursts. Twersky [8] and Witcher [9] each produced useful instruments of this type.

A travel-aid instrument should employ the native ability of the user to the maximum possible extent; it should not try to replace functions that the human being already possesses. Thus, at first glance, an instrument using sound might seem to be very desirable. However, it is not possible to obtain a sufficiently narrow sonic beam for frequencies in the audible range. The highest frequency that is practical seems to be 10 kc. With frequencies higher than this, too many people are unable to hear the reflections. Although under ideal conditions the device works well, under actual operating conditions it makes enough noise to attract attention, and the beam width is so wide that reflection from the ground confuses the pattern.

There is one factor in favor of using ultrasonics in guidance devices.

The rather slow speed of transmission of ultrasonic waves makes it possible to employ modified-radar principles. This is not possible with electromagnetic waves, since the distances are too short and the speed of transmission is too high. However, none of the ultrasonic devices built[1] have proved satisfactory, since these devices use air as the conducting medium, and thermal and convection currents introduce refraction effects that frequently completely obliterate signals. In fact, one evaluator has vividly recalled the disturbing experience of standing still and observing a tree trunk about 10 ft away come and go.

In addition, reflection using feasible frequencies is specular (mirror-like), and thus it is necessary to have the incident beam of energy strike the surface to be detected at right angles. This, of course, is not practical. Specularity results because most surfaces are too smooth for waves of approximately 1 cm in length. Yet if the wavelength were reduced to the point where surfaces scattered sufficiently, absorption by the air would make the amount of power required prohibitively high.

There are also some purely mechanical difficulties. Obtaining a sufficiently narrow beam of ultrasonic energy requires the use of parabolic reflectors with apertures of twenty to thirty times the wavelength. This causes the apparatus to be too large. Two such reflectors would be necessary for each device, one to form the outgoing beam and the other to gather the returning energy for the receiver. Furthermore, the ultrasonic generator is somewhat cumbersome. An oscillator and a transducer are required, in contrast to the simple lamp required for an optical system, for example. The transducer in the receiving system is also inconvenient for ultrasonic waves.

Consideration of all these difficulties leads to the conclusion that in the present state of the art ultrasonic waves can be used only for a simple go, no-go obstacle detector and that they cannot be used to detect such fine details as curbs. Moreover, even obstacle detection can presently be done more simply and more compactly with devices using light.

Another possible approach is the use of radar. A true radar system would be too heavy for portable use and unsuited to the very short ranges involved in a practical device for the blind. However, one useful system has been suggested; it comprises a $\frac{1}{4}$-wave antenna that changes the frequency of a high-frequency radio oscillator when it nears an obstacle. This is the same principle that is used in the proximity fuse [3].

Three interesting devices have been developed using electromagnetic radiation in the infrared and the visible regions. One of these is an extremely ingenious and well-engineered instrument called Optar [4].

[1] The Hoover Company, The Brush Development Company, and Stromberg-Carlson have built such devices. For further information see Roberts [5] and Slaymaker and Meeker [6].

Detecting ambient light reflected from the obstacle, it determines the range by automatically locating the distance behind a lens at which the image of the obstacle is in sharp focus. Information is presented as a tone, the pitch of which indicates distance. In practice, the tone is complex and difficult to interpret, and the device in the form in which it was constructed is not capable of detecting down curbs.

A second device, developed by the Signal Corps, works in the same part of the spectrum but uses optical triangulation.[2] Twenty-five experimental units have been manufactured. The principle of operation is as follows: A beam of light, interrupted 500 times per second, emanates in a narrow beam from an optical system, strikes an object or the ground, and is reflected back into a second optical system. This second system focuses the image of the spot of light reflected from the object on a coding disk. The coding disk, according to the distance to the object, interrupts the reflected light 4, 8, 16, or 32 times per second. Behind this disk is a photoelectric cell whose output is fed through an electronic amplifier tuned to 500 cycles and then to a vibrator in the handle of the instrument. The 500-cycle pulses from the vibrator in the handle inform the user of the distance to an obstacle or discontinuity in the terrain; the position of the handle indicates the azimuth, i.e., whether the obstacle is to the left, to the right, or straight ahead.

A third optical device, using triangulation and visible light, was built by Witcher at the Massachusetts Institute of Technology. He gave considerable time and thought to the overall problem of obstacle detection, the special aspects of auditory localization, the presentation of information in a maplike display (extending the work of Sokal [7]), and the detection of step-downs or curbs. For this last problem, he developed a test model of a device, intended to be fitted in a briefcase, that projected a beam of light obliquely forward and downward upon the ground and then scanned this beam forward and backward along the ground in front of the user. The beam was projected through the lower of two lenses, and a photocell was placed behind the upper lens. So long as the spot was on level or uniformly sloping ground, the optical system received a solid, uninterrupted spot of light. If, however, the beam of light projected over the edge of a curb, there was a momentary interruption in reception because the sudden change in distance between curb and street caused the spot to "jump" and thus disappear momentarily from the range of the optical system. There seemed to be some difficulty in keeping the optical system in focus over the rather wide range of distances involved; however, this difficulty probably could be overcome by adding an automatic focusing arrangement coordinated with the

[2] Under the direction of Lawrence C. Cranberg. See Cranberg [2].

back-and-forth scanning of the spot. Dr. Witcher's untimely death prevented further development of this system.

Under the National Academy of Science's Committee on Sensory Devices, the Franklin Institute Laboratories for Research and Development worked on an optical-triangulation ranging system using ultraviolet light. Ultimately, however, the basic idea was given up because of numerous technical difficulties attendant upon the use of ultraviolet light.

In 1953 the Veterans Administration contracted Haverford College to develop, under the direction of the author, an improved version of the previously mentioned Signal Corps device. Haverford subcontracted the laboratory development to Biophysical Instruments, Inc., leaving its evaluation to Haverford. A version of the original Signal Corps device had been evaluated by the author in 1950, and the resulting report of that evaluation [1] contained a list of recommendations that ultimately became the specifications for the new instrument to be developed by Haverford.

The development project resulted in the production of three prototype obstacle detectors as well as a portable study model of a curb detector. Reproduced below are the recommendations cited above concerning the original Signal Corps device, together with annotations of the success achieved in implementing these recommendations to date with the new instruments.

Desirable Improvements in Order of Importance	**Action Taken**
1. Separate obstacle and curb locators that deliver no signal until needed and that are not accidentally actuated by irregular movements of the instrument encountered during walking.	1. Achieved on obstacle detector (OD) and on curb detector (CD) as far as it is developed.
2. Two-channel presentation; one for curb signals, the other for obstacle signals. The stimuli might be applied to the user through small diaphragms, with one channel for the index finger of the hand and the other channel for the middle finger.	2. Presently one-channel obstacle-detector presentation. Another channel is reserved for curb detector. Best frequency found to be 20 cps with the poking stimulator.
3. Automatic scanning for obstacle detector, with a path about 3 ft wide at a distance of 8 ft.	3. Manual scanning adopted for sake of simplicity.

Desirable Improvements in Order of Importance	Action Taken
4. Extreme simplicity in getting instrument into operation. The user would prefer simply to pick up the device, "flip a switch," and be ready to travel.	4. Achieved.
5. Include feature to aid the user in walking a straight line.	5. Not investigated. It is considered, however, that use of a tuning-fork-type stabilizer[3] would make a good starting point.
6. Quieter motor. Vacuum-tube or transistor switching might be employed to eliminate need for motor, chopper, and coding disk.	6. All moving parts eliminated on OD. CD presently uses a chopping disk and a servomotor. Thought has been given to ways of eliminating these moving parts on an actual model.
7. Facility for rendering beam invisible when desired.	7. Beam has purple glow, but light image can be perceived only in dark surroundings.
8. A dual-range control for obstacle detector at either 10 or 20 ft.	8. Dual-range control with ranges centering at 4 and 7 ft.
9. Provide a guard that will prevent women's skirts, dresses, and overcoats from getting in front of lenses.	9. No longer considered necessary.
10. Straight handle to facilitate horizontal positioning of the instrument. The handle in the model tested was curved intentionally to aid the user in keeping the spot of light on the ground at a given distance away. If item 1 is achieved, horizontal positioning will be important.	10. OD has straight handle. CD design is such that horizontal positioning is unnecessary because of pendulum-type mount.
11. Distance discrimination to within 1 ft of instrument.	11. Present dual-range detector operates to within 2 ft.
12. Attempt to reduce "noise" (background interference) due to bright sunlight.	12. No background noise except for single pulse from sudden bright-sunlight reflection.
13. Reduce weight; storage-battery life of 4 hr is sufficient.	13. Weight of OD reduced from $4\frac{1}{4}$ to 2 lb, with storage-battery life of 10 hr. Could reduce weight somewhat more if necessary.
14. Provision for carrying the instrument around the neck. This will require a different type of stimulator. This item is mentioned to stimulate	14. Not investigated.

[3] Manufactured by Sperry Gyroscope Company.

Desirable Improvements in Order of Importance	Action Taken
thought in order to improve the instrument so as to free both hands of the user.	
15. Logarithmic response in amplifier to reduce variations in intensity of received signal that are due to varying reflectivity of objects.	15. Stimulator intensity is independent of received signal intensity.
16. Progressively complex models to accommodate users of varying degrees of ability.	16. Not yet relevant.
17. Keep acoustical noise of tactile stimulator below point at which it can be heard under normal travel conditions.	17. Achieved.

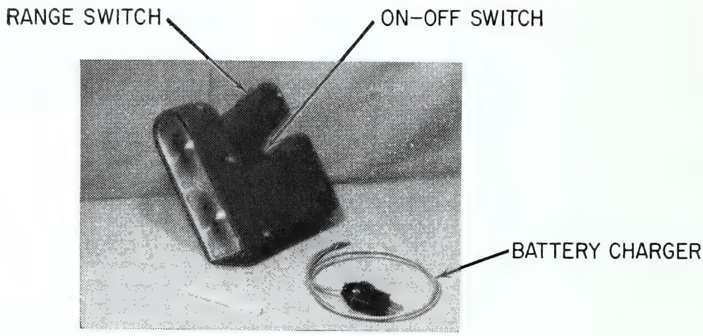
Figure 22-1 presents two photographic views of the dual-range obstacle detector. A diagram of the principal parts of the device is presented in Figure 22-2.

The action of the detector starts with a pulse generator. The pulses are amplified and used to excite a xenon lamp. The resulting light flashes are of the same duration and frequency as the initiating pulses. The lower lens focuses the light into a beam. Reflections from obstacles in the beam are caught by the upper lens and focused onto a pair of photodiodes. The lower, called the far diode, is mounted at the focal point for reflections from distant obstacles. The upper, or near, diode is at the focal point for reflections from close obstacles. When the range switch is pressed, the far diode is cut out; at other times both diodes are active.

The photodiode currents are amplified by the light-signal amplifier. The output of the pulse generator is amplified by the gate-signal amplifier. This amplified pulse is used to open a gate at the correct moment for receiving any pulse from the light-signal amplifier that might result from a reflection of a flash from the xenon lamp. The gate is closed at all other times in order to discriminate against spurious light from other sources. Sufficiently large light-signal pulses passed by the gate will trip the one-shot multivibrator. When triggered, this circuit produces a pulse of sufficient magnitude to drive the stimulator, causing it to give the finger a single poke.

Other parts that should be enumerated are: the power supply, which includes batteries and an on-off switch; a socket for receiving a charging current; a chronistor that measures by electrolytic deposition the amount

of use that the instrument has been given; a jack for external monitoring of the stimulator signal (useful for demonstrations and for training, to allow the trainer to know what signal the trainee has received and thus to know whether a lack of response is the fault of the instrument or the trainee); the case that houses all the previously listed parts; and a



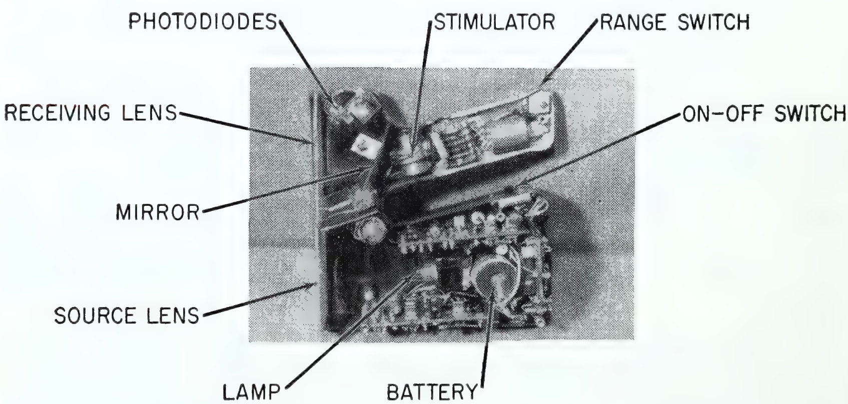


FIG. 22-1 Two-range obstacle detector.

storage case, which contains a built-in battery charger and a supply of Drierite for removing any moisture that may get into the instrument when used in damp weather.

The case of the instrument itself is molded of Fiberglas-reinforced epoxy. It is a two-piece shell with grooves to hold the plastic Fresnel-lens windows. There is a small door that can be opened for access to the lamp and to several test points without the need for removing the whole side. Several sections of the inside of the case are painted with

conducting paint and grounded for shielding purposes. A ring is provided for fastening a carrying strap.

After the three dual-range obstacle detectors were completed late in 1958, an evaluation study was begun. For the first few months numerous difficulties were experienced in keeping the devices in operation. They were used by 13 subjects for an estimated 150 travel periods, each lasting approximately 1 hr. As a result of these tests, a carrying strap was provided for people who must also carry books, pocketbooks, and packages. The handle size was tested by eight women. Four of them found the handle slightly large but not uncomfortable, and four found it to be acceptable. All women found the distance between the

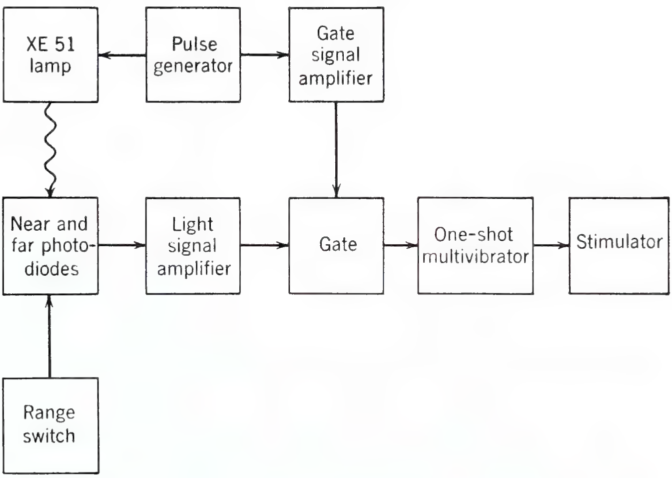


FIG. 22-2 *Block diagram of dual-range obstacle detector.*

range switch on top of the handle and the stimulator under the handle to be uncomfortably great. Four of the women found it impossible to sustain a grip on the range switch. It was noted that occasionally the device did not detect rain-soaked obstacles effectively. Tests were also made with the device at temperatures of 25°F and below, with the device left outdoors for 5 hr or more. In every case at least ½ hr of warmup indoors was required to make it operative.

Certain effects have been observed with the obstacle detector for which no immediate explanation is apparent. Upon approaching an obstacle, for example, the user will be signaled at a given distance. If he then backs away from the obstacle, he will continue to receive the warning signal for distances considerably greater than that at which the detection originally took place. Some objects, such as some brick and cinder-block surfaces, windows, and glass doors, certain dark cloth

surfaces, small poles, and objects of open-wire construction, were found difficult, and in some cases impossible, to detect.

In the early stages of training it was difficult to impress upon the trainee that this device would not, without the curb detector, ease the strain of traveling in strange territory. However, once the user had become accustomed to the device, his walking speed and confidence improved greatly as training progressed. It is necessary early in training to teach the subject to subdue his curiosity and concentrate on his walking. The urge to identify objects that the detector has signaled must be changed to the simple purpose of missing them. Most of the trainees were surprised to find that the device does not take the place of aural and kinesthetic cues. Two people found it so difficult to synthesize their previous means of guidance with the new information gained from the device that they failed to progress in their training beyond the elementary stages. Others found the device a burden at first, but most, with practice, discovered its advantages, realized its limitations, and progressed well, using other cues to supplement the information given to them by the detector.

Though the present obstacle detector can detect almost all obstacles encountered, the number missed reduces the confidence of the user. However, there are sufficient travel situations in which the device is useful to make it clear that an obstacle detector, even without a curb detector, is a very important instrument. It is useful primarily in semifamiliar territory in which the user already knows in a general way where the corners, steps, and curbs lie but cannot be sure of their exact location within one or two steps and must contend with movable objects that are in unpredictable locations. In semifamiliar environments the dual-range detector, when pointed at the ground, serves effectively in locating extreme changes in terrain, such as the edge of a subway platform or the top of a flight of stairs, the existence of which is known but the exact location of which must be determined.

The obstacle detector is also of considerable use in crowded buildings. One of the trainees was a blind high-school student in a sighted school. He said that, without the device, going to school was sheer torture. Whenever he had to change classes, he bumped into doorways and tripped students with his cane. The corridors were so noisy and crowded that aural cues were useless. With the detector he was able to walk at normal speed and rarely missed a doorway or bumped into a fellow student. After the first day with the obstacle detector, he abandoned his cane for the duration of the field test. However, he did not have confidence in the obstacle detector outside, for it could not detect curbs.

The purpose of a step and hole detector is to give warning of an

abrupt discontinuity in the terrain. This warning must be given at sufficient distance to allow the user time to modify his course. The discontinuity may be either an up step or a down step and may take any of several forms, e.g., curbs and projecting stairs on building entries. The term break is used to designate any such discontinuity.

As with the obstacle detector, it is considered important to keep the ears free for auditory cues by means of using a tactile stimulator that presents a signal only when necessary and is silent at all other times in order not to fatigue the user. Indication of range or azimuth is not considered important, but the instrument should be designed to indicate the location of an object as close to a fixed range as practicable, and the solid angle of view should be as small as possible to allow the user to scan manually.

Only three pieces of information need be presented: (1) an up step; (2) a down step; (3) no light return, which the user would have to interpret as a hole, a type of surface that the instrument cannot detect, incorrect carrying of the instrument, or mechanical failure. Since failure of the instrument gives the same signal as a curb or hole, the operation of the device is known as fail-safe; that is, the user is warned and stops to investigate.

It is necessary that at least one step be allowed after warning of a break for the user to react to the warning, shift his weight, and decide how to alter his course (see Figure 22-3). Break detection can take

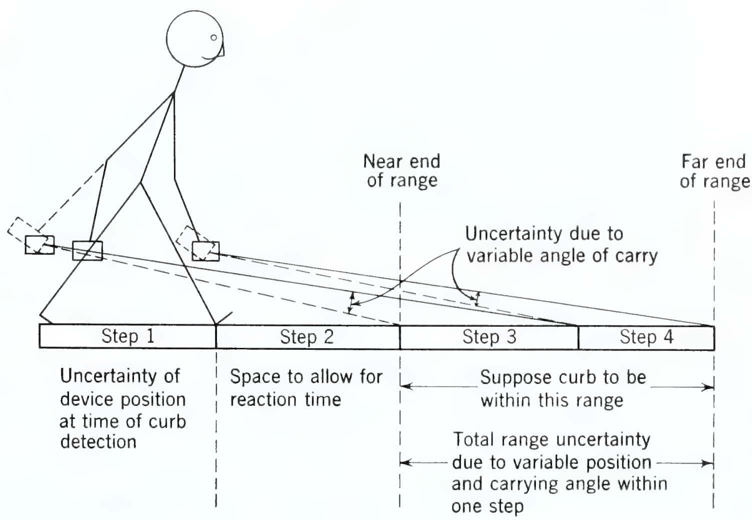


FIG. 22-3 *Illustration of the considerations and assumptions involved in determining the range within which a curb should be detectable.*

place at any time during a step, and one more step must be allowed for reaction time. Normal step length during walking is 30 in. Thus the minimum detection distance should be about 60 in., so that even long-striding individuals may have one step in which to adjust. When the device is used for static sensing (i.e., locating a break while the user is standing still), a shorter range might be preferred. Therefore, 60 in. is probably the best design center, and the range over which a break should be detectable is 3 to 7 ft.

There are several points on which all break-detection systems should be judged:

1. Insensitivity to tilting by the hand
2. Insensitivity to slope of terrain
3. Insensitivity to change in height from ground
4. Power required for adequate detection sensitivity under all conditions of operation
5. Inherent minimum size established by optical system
6. Detection of curb of a minimum height of 2 in. at a distance of 3 to 7 ft
7. Total weight in combination with obstacle detector (must be less than 5 lb)

With one exception, all the many systems reviewed over a long period of time were found to fail to satisfy requirements in one or more of these areas.

Two models of the curb detector have been built: a bench model and a portable study model. The bench model was set up to study the optical and servo systems and the circuitry. The portable model is a mechanical unit mounted on bearings that is used to test the oscillation propositions developed and to enable a person to carry it in a field test. A photograph of the portable unit is presented in Figure 22-4. This model was not intended to be compact enough and adequately enough housed to be a practical working instrument. However, it was hoped that it would yield enough performance data to bring the design of a practical instrument within sight. It has not yet been tested fully, and therefore details of its performance are not available.

Figure 22-5 is a block diagram of the curb detector. The main beam-forming lamp, shown in the upper right-hand corner of the diagram, is a miniature incandescent lamp with a coiled filament covering an area of approximately ½ by ¼ mm. The light is chopped by a chopping wheel at a rate of 1,200 cps. The light and dark portions of the cycle are of equal length, and the pulses of light are roughly rectangular in waveform.

The light beam is formed by a Fresnel lens of 4 in. diameter and 3 in. focal length. As the device is ordinarily carried, this beam leaves the lens about 30 in. above the ground and 10 in. in front of the hand. It hits the ground about 5 ft in front of the hand. The diameter of the spot on the ground is ½ to ¾ in. The light from an auxiliary lamp is

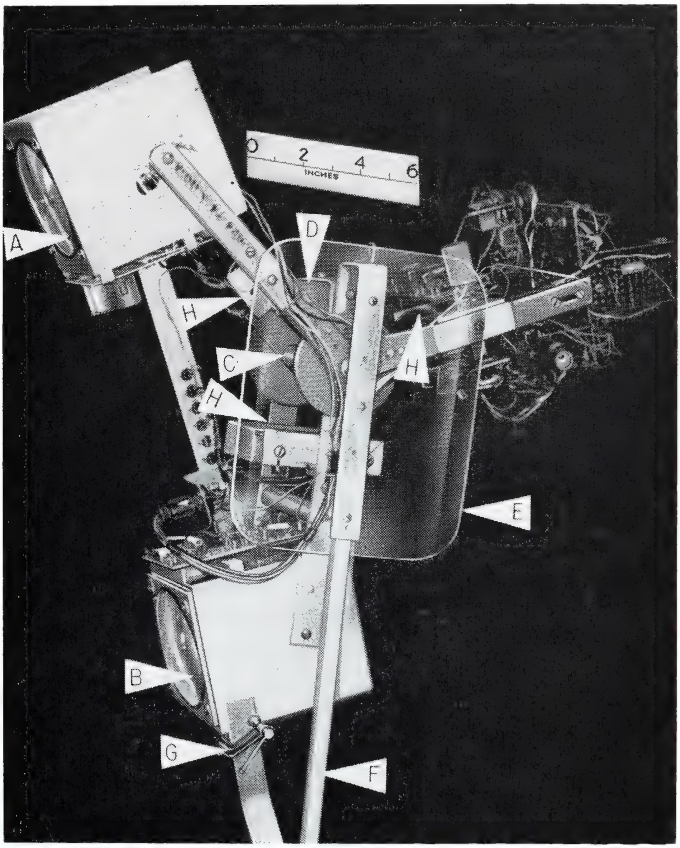


FIG. 22-4 *Portable curb detector. (A) Source lens; (B) receiver lens; (C) ball bearing; (D) auxiliary handle; (E) body and clothes shield; (F) auxiliary cane; (G) clamp to stop swinging; (H) batteries.*

chopped by the same wheel and then enters a photocell, the input of which becomes the gate signal. It is amplified by the gate-signal amplifier and is used to drive the two gates.

At the lower right of the figure is shown the receiving lens. It is identical with and mounted about 15 in. below the source lens. It picks up light scattered from the target and focuses it on a pair of photocells. These photocells are mounted on the armature of a servomotor, which

moves them in order to follow variations in position of the light spot. As presently adjusted, the range limits between which the cells are able to track the spot are 40 and 87 in., as measured from source lens to target.

The tracking is accomplished by noting any intensity differences between the two photocells' responses and feeding the servomotor a signal that will cause it to move in such a direction as to decrease the difference. The signals from the photocells are amplified in separate channels, labeled *A* and *B* in the figure. The amplified signals pass through

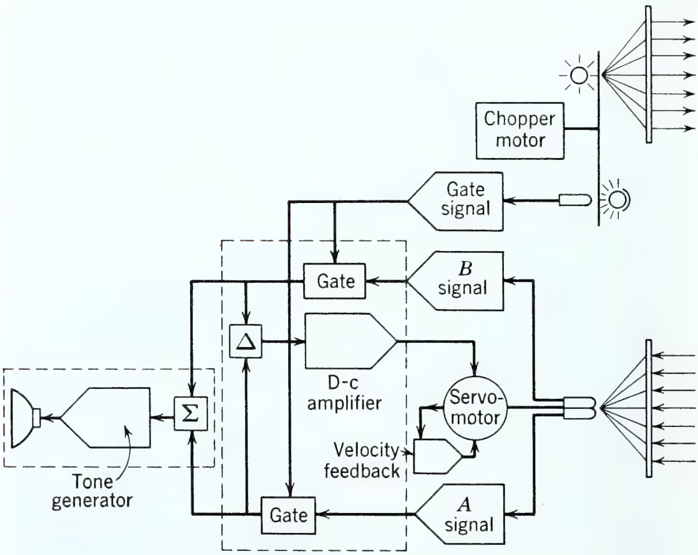


FIG. 22-5 *Curb-detector block diagram.*

the gates, which are synchronized by the chopper so as to discriminate against any signal that does not have the frequency and phase of the light in the instrument's own beam. The outputs from the gate are fed into two circuits. One circuit is the sum of the two signals and is labeled Σ; the other is the difference between the two signals and is labeled Δ. The difference signal is amplified and is used to drive the servomotor so that it will track the light spot as described above.

The figure shows in connection with the servomotor an amplifier labeled velocity feedback. A small magnet mounted on the servomotor armature induces a voltage in stationary pickup coils whenever the armature moves. This voltage is proportional to the angular velocity of the armature. The velocity-feedback amplifier amplifies this voltage and applies to the servomotor a torque that prevents increase in velocity.

As a result, the velocity, rather than the acceleration of the servomotor, is made proportional to the photocell signals driving the motor. This helps to prevent oscillations of the servomotor and may also prevent it from following rapid changes in positions of the light spot such as might be produced by encounter with a curb.

The sum signal is fed to a tone generator. This circuit responds only when the sum of the two light signals falls below a predetermined value, which may be fixed by a potentiometer setting. Whenever the sum falls below this value, i.e., whenever the photocells lose the light spot, the circuit responds by emitting a squeal through a miniature speaker. In a practical model the squeal would be replaced with a stimulator for poking the finger, as in the obstacle detector, or with some other device that does not compete for the attention of the ears.

The portable-model curb detector has been carried by 10 or more people and has consistently detected up and down curbs, boxes, walls, and other objects that are encountered by its downward-slanting beam. It always detects down curbs of about 3 in. or more and up curbs of about 6 in. or more. It tracks well over most surfaces, including mud puddles, but has some trouble on an asphalt pavement. The presence of bright sunlight was troublesome in some of the outdoor trials.

During the testing, the device was carried both swinging freely on its ball bearings with a period of about 10 sec and mounted on an auxiliary cane as shown in Figure 22-4. It detected curbs equally well with both arrangements. However, performance in other respects was definitely superior when the detector was clamped. When the device swung freely, three annoying characteristics resulted:

1. The detector was always swinging out of the range over which the servo could track the beam. Whenever the light spot swung too far out or too close, a false squeal resulted. It was constantly necessary for the carrier to use his other hand to steady the device and to bring it back to a workable position.

2. One could never be sure how far away the light spot was. When a curb or obstacle was detected, the spot might be anywhere from 3 to 5 ft distant.

3. When the beam was lost by the servo, it was necessary for the carrier to use his other hand to tilt the device forward in order to regain the beam. When the device is clamped to the handle, one has a feeling of better control. One can tilt the device to choose one's range. If the beam is lost by the servo, it is necessary only to tilt the device forward to regain the beam.

Why these preliminary results do not seem to bear out the prediction that the free-swinging mount would be necessary is one of the problems that needs to be studied. It may turn out that extra steadiness is required

for detecting up curbs lower than 6 in. If so, the preliminary trials indicate that further refinements are necessary to eliminate the annoyances that now accompany the free-swinging mode of operation. It may also turn out that the excessive weight of the device (9 lb) prevents the user from moving it freely and that even a 4- or 5-lb device would present a mobility problem.

On the basis of the research, the following areas for development are recommended.

Obstacle Detector

1. More rigid optical system
2. More rugged mechanical design of the circuits
3. More convenience of access for servicing
4. Weatherproofing
5. Relocation of the range switch
6. Repositioning of the stimulator
7. Trial of the new xenon FX-6A lamp[4]
8. Spherical mirror behind lamp to increase light output
9. Addition of a third photodiode in the extra-near position (parallel with the present near diode) to extend the near range to closer distances
10. Setting of the discriminator in the present gate circuit to reject all pulses smaller than the largest noise pulse likely to be encountered often
11. Provision of a servicing socket containing a large number of terminals leading to all important test points

Curb Detector

1. Thorough study of the present model to determine the size of the up-curb signal available from the velocity-feedback circuit and the possibilities for detecting smaller up curbs
2. Consideration of the feasibility of adding an up-curb signal generator
3. Addition of a beam-finding circuit
4. Possible substitution of the FX-6A flash tube for the present lamp and chopping motor
5. Reduction of the diameter and focal length of the sending lens unless a broader light-source area is used
6. Possible replacement of the mechanical servo system with an array of fixed photocells served by an automatic switching circuit

[4] Developed by Edgerton, Germeshausen & Grier Co.

7. Study of reaction to bright sunlight

8. Reduction of the bandwidth of the amplifier or use of full- instead of half-wave gates to discriminate between d-c and low-frequency components

9. Establishment of a stimulator with a poking rate proportional to nearness of the objective

10. Use, for obstacle detection, of that portion of light emitted by the source that is at right angles to the rays usable by the curb detector

11. Study of the reaction to asphalt pavements

12. Consideration of a design incorporating most of the components of the present model, assuming that further tests do not encourage radical changes (such a design is shown in Figure 22-6)

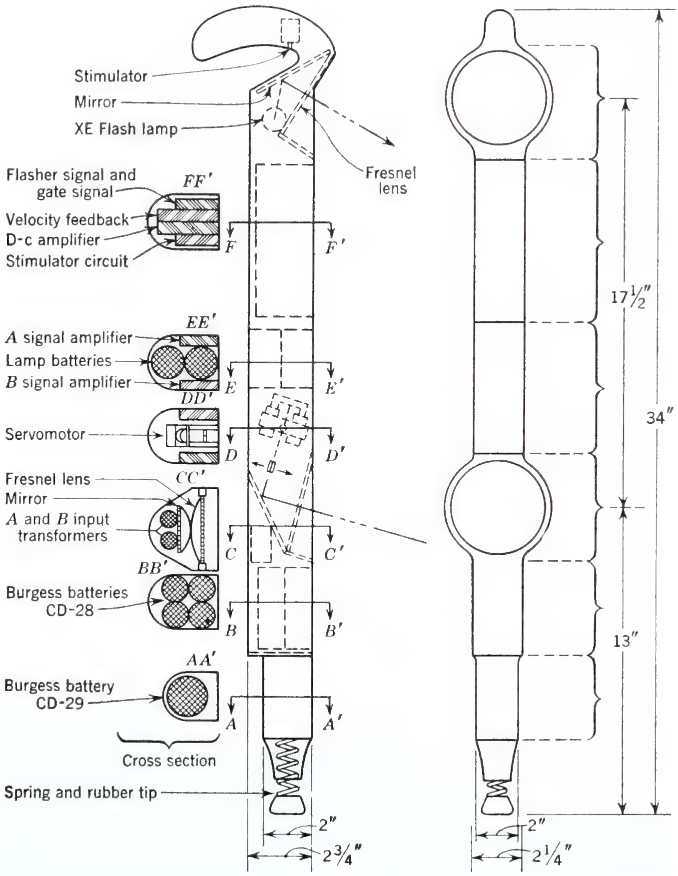


FIG. 22-6 *Possible design for the next curb detector. The braces at the right of the figure indicate the front panels, which are removable. For easy access the inside parts of the curb detector are attached to these panels.*

BIBLIOGRAPHY

1. Benham, Thomas A.: Evaluation of Signal Corps Sensory Aid for the Blind, AN/PVQ (XE-2), Contract V1001M-1900, Apr. 25, 1952. (Copies may be obtained from the author.)

2. Cranberg, Lawrence C.: Sensory Aid for the Blind, *Electronics,* vol. 19, no. 3, pp. 116–120, March, 1946.

3. Freedman, Samuel: Midar, *Radio Television News,* vol. 42, pp. 30, 109–111, August, 1949.

4. Kallman, H. E.: Optar, A Method of Optical Automatic Ranging as Applied to a Guidance Device for the Blind, *Proc. IRE,* vol. 42, pp. 1438–1446, 1954.

5. Roberts, A.: Ultrasonic Echo Sounding Equipment for the Blind, *Radio-Electronic Eng.,* vol. 12, no. 6, pp. 6–8, 28, June, 1949.

6. Slaymaker, F., and W. F. Meeker: Blind Guidance by Ultrasonics, *Electronics,* vol. 21, pp. 76–80, May, 1948.

7. Sokal, N. O.: Area Display of Obstacle Location for Use with a Guidance Device for the Blind, master's thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, Mass., September, 1950.

8. Twersky, V.: Auxiliary Mechanical Sound Sources for Obstacle Perception by Audition, *J. Acoust. Soc. Am.,* vol. 25, no. 1, pp. 156–157, January, 1953.

9. Witcher, C. M., and L. Washington, Jr.: Echo Location for the Blind, *Electronics,* vol. 27, no. 12, pp. 136–137, December, 1954.

10. Zahl, Paul A. (ed.): "Blindness, Modern Approaches to the Unseen Environment," Princeton University Press, Princeton, N.J., 1950.

Chapter 23

THE FRANKLIN INSTITUTE ELECTRONIC CANE

*Robert J. Gibson**

A CANE THAT POSSESSES THE NORMAL FUNCTIONS OF A BLIND MAN'S STICK but is enhanced through electronics by the additional ability of warning its user of step-downs in his path has been under development and evaluation since 1947. It is believed that there are four factors that influence the success of an electronic travel aid.

First, despite the fact that many devices have recently been developed for the blind traveler, the primitive cane—a simple extension of the arm—is still the most successful and popular guidance device. This militates against any concept that departs far from the ancient blind man's stick.

Second, the consensus of advice, written and oral, from the blind themselves and from those who work with the blind is that the detection of a step-down such as a curb or manhole is a most vital piece of information to the blind traveler. In addition, the greatest number of travelers can absorb and use only a single piece of extra guidance information. If two or more pieces of information are supplied to him, confusion is likely to result. Training him to respond to one type of stimulus is simple and rapid.

Third, the user should receive a warning stimulus only when, and for as long as, the step-down hazard threatens him. The argument for this point is as follows: Often a continuous stimulus fed to the traveler either is a hindrance or is ignored. A single strong stimulus, a warning of danger, is usually met immediately with the appropriate avoiding reaction. In most cases, all the traveler's operative senses would be alert and feed to his brain whatever cues of the environment they detect. There is no question that a continuous stimulus in one sensory channel

* Bio-Instrumentation Laboratory, The Franklin Institute Laboratories for Research and Development, Philadelphia, Pa.

The program reported in this chapter was supported by the W. K. Kellogg Foundation and by the Office of Vocational Rehabilitation of the U.S. Department of Health, Education and Welfare.

would inhibit other information in that channel and often interfere with information in other sensory channels as well.

Fourth, a travel aid must be fail-safe. If for any reason the guidance device should fail or the operator become confused by the stimulus, the cane would still be a cane and could be used as such. As simple as this notion is, it is most important, for in crowds, in vehicles, on stairs, or in confined quarters the cane is primarily a probe, an extension of the arm, and is unchanged and unrivaled as an obstacle detector.

These criteria were the guideposts during the development of the guidance aid described in this chapter.

CONFIGURATION

The electronic cane discussed here is shown in use in Figure 23-1. It has a hollow, tapered shaft approximately 3 ft long and is made of

FIG. 23-1 *Electronic cane in use.*

glass fiber bonded with plastic. At the end of the shaft there is a replaceable hardened-metal tip. The handle is made of impact-resistant white plastic and contains all the electronic equipment. Figure 23-2

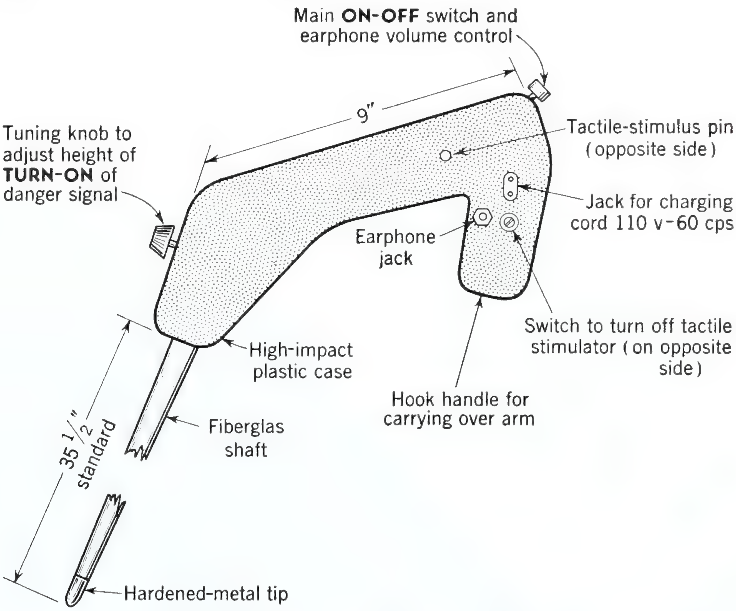


FIG. 23-2 *Exterior of electronic-cane handle.*

shows the outside of the handle, with the important details indicated. Figure 23-3 shows the position of the electronic equipment in the opened handle. At the forward end of the handle is a small knob for tuning the electronic circuits for the height at which the tip of the cane would normally be carried. At the rear is a knob for turning the circuits off and for adjusting the volume of the audible signal. The cane is gripped as in holding a suitcase with the handle horizontal. A small plastic button protrudes slightly from the right side of the handle. When the tip of the cane passes across the edge of a step-down, the button signals the user by vibrating against his hand. In addition to this signal, the circuits generate a steady whistle that can be heard in a hearing-aid earpiece plugged into the cane handle.

USE

The chief advantage of the electronic cane is that it sweeps continuously over the path of the traveler, whereas a tapping stick samples small bits of the path and only at the points touched by the cane tip. Tapping is unnecessary with the electronic cane.

The electronic cane gives a stop signal when the tip passes over a down step or curb or over the edge of a hole. When a falloff of the ground level is detected by the electronic circuit and the signal to stop is given, the shaft should then be used to probe the depression in a normal fashion.

The cane is held in the hand with the arm relaxed and extended slightly forward. When the cane is held at its balance point, its shaft assumes a natural position across the body, with the tip in front of the

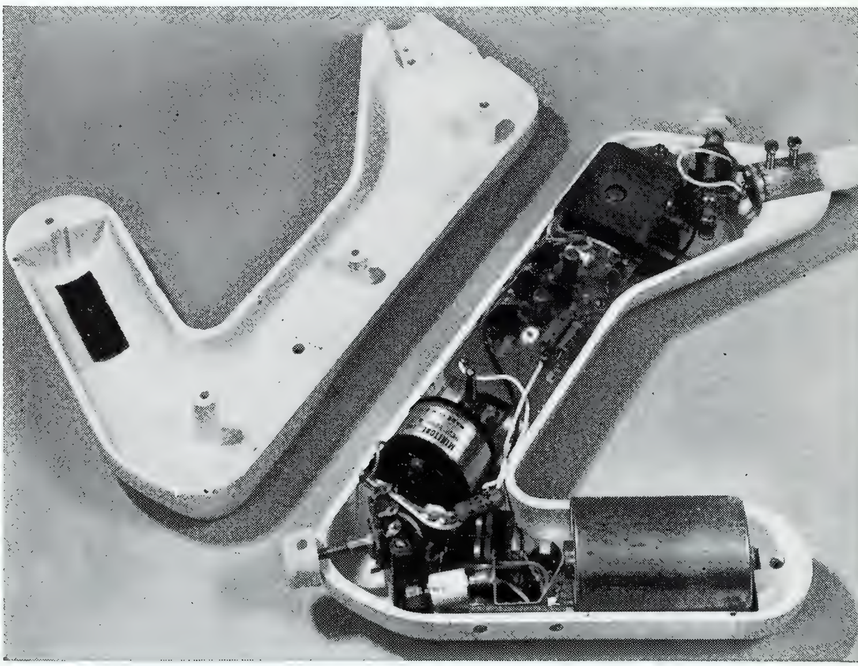


FIG. 23-3 *Interior of electronic-cane handle.*

left foot and at the proper angle to the ground. By twisting the arm slightly, the user can sweep the tip of the cane across the path to a point in front of the left foot as the right foot advances. The tip is easily swept in a plane parallel to and approximately 2 in. above the ground, thus avoiding a rising or dipping motion. At first, this manner of holding and sweeping the cane may cause some fatigue. However, it is ordinarily a comfortable position, and the sweep is only slightly different from some recognized methods of conventional cane technique. The level sweep can be learned rapidly. In use over long periods of time, the electronic cane should, because of its balanced construction, cause less fatigue than a standard cane; although the electronic cane is heavier than a wooden cane, its balance and large, comfortable handle make it easy to carry.

When they first try the cane, most users find that false signals occur frequently. A false signal is defined as a signal caused by the tip's being raised inadvertently above the set altitude. The cane does not under normal conditions give a signal unless the distance from the tip to the ground becomes greater than the set distance. With a little practice the user can find the best height for the set point; 2 in. is about the average distance. At this height some leeway is allowed for rises and dips during the sweep. The most frequently occurring error is a sweep that is high at both ends or low at both ends.

Tuning of the signal is accomplished by raising the tip 2 in. from the ground and rotating the tuning knob just far enough to cause the signal to stop. Lifting the cane tip an additional $1\frac{1}{2}$ in. should cause the signal to start. Tuning soon becomes second nature to the user, much as one tunes a radio without thinking. At any time during a walk, the user can reassure himself that the cane is operating by simply lifting the tip for a moment and listening for the signal. It is recommended that this be done frequently during the early part of training, although the numerous false signals caused by a novice's handling will have the same effect.

As has been mentioned, the signal to stop may be a vibration on the hand, a squeal in the earphone, or both. The beginner is urged to use both signals. Further, if the audible signal is used during teaching sessions, the instructor will be able to hear it and correct bad habits of swing that cause false signals or determine whether or not the student pays attention to the stop signal when he should.

The response of the cane to a dropoff is very quick. Normally the tip of the cane will have passed only a few inches over the dropoff before the stop signal occurs. Thus a warning of at least one step is given to the user.

CIRCUITRY

The electronic cane is essentially a proximity device capable of detecting a distance greater than 2 in. from any grounded surface, including asphalt, concrete, sand, and wooden boards on the ground. Its circuit contains two oscillators. One is in a tuned circuit that contains a capacitor whose plates are a conductor at the tip of the cane shaft and the earth. The other is in a tuned circuit that contains a small variable capacitor controlled by the tuning knob at the front of the cane's handle. This tuning compensates for changes in body capacity that can be caused by wearing gloves or various types of footgear.

As the distance between cane tip and ground increases, the capacitance of the capacitor in the first oscillator circuit decreases, thereby

changing its frequency. The two tuned circuits or oscillators, operating at approximately 2 Mc, are beaten against one another and locked in sufficiently to provide a sharp onset of any difference signal. Thus, once the fixed oscillator has been tuned to the frequency of the variable oscillator by holding the tip at a given height above ground, a difference signal occurs when the ground falls away from the tip. Detected by a diode detector and amplified, this difference signal provides the signal heard in the earpiece. Further amplified and smoothed, the difference signal operates a small motor that in turn drives an eccentric that is attached to the plastic pin to give the tactile stimulus.

A built-in 110-volt charging circuit with separate line cord serves to recharge the 10.8-volt nickel-cadmium cell overnight, storing enough power for a full day's use by an experienced traveler. The cane may be left on charge indefinitely without harm. The current drain on the battery is less than 8 ma with the motor not operating, rising to between 100 and 200 ma with the motor on. Since this high drain is intermittent (lasting usually 1 to 5 sec), there is no problem of excessive power drainage.

DESIGN FACTORS

Early in the development of the cane it was realized that an auditory stimulus might interfere with other auditory cues. However, it was not until recently that it became possible to provide the power necessary for making a tactile stimulus equal to the stimulus of an auditory cue. A tactile stimulus was necessary for three reasons: first, acoustic shielding of the ear is caused by the earpiece; second, many of the blind object to the inconvenience and appearance of an earpiece; last, the necessity for providing a stimulus that would not be masked by heavy traffic noises.

The tactile stimulus had to be powerful, requiring 50 to 100 times more power than an auditory stimulus. The literature was of little use in this problem, since most of the information concerned just-noticeable stimuli, optimum frequency at low levels only, most sensitive areas of stimulation, and similarly irrelevant factors. A stimulus had to be produced that would always be noticed, even through gloves, and that would not require careful and fixed placement of the hand. This could not be done with less than about 0.5 watt of power input to a reasonably efficient electromechanical device. A small motor with an eccentric pin ($\frac{5}{32}$-in. diameter, rounded end) giving a $\frac{3}{32}$-in. maximum stroke through a small hole in the side of the handle was found suitable. The motor when operating uses from 0.5 to 1.5 watts for a few seconds, depending on how tightly the hand is pressed against the pin. A strong stimulus is obtained even with a strong grip and on any part of the hand

that touches the pin. The location of the pin is such that it is difficult to place the hand so that the pin does not touch it at some point.

The acoustic signal was retained because of its value to the instructor in the training program.

Another factor in the design of the cane was the stability of the set point of detection. Although the set point could be reset at any time, drift could be troublesome and even dangerous. If the traveler were carrying a package, for instance, he could not retune the cane without difficulty. For some time in the evolution of the cane, temperature change was a source of drift of the set point. Temperature stability was finally achieved by making the two oscillators identical, including the cable down the shaft. In this way any drift in frequency of the one oscillator is compensated by a similar drift in the other. In present canes retuning is unnecessary when on a cold day the cane is taken outside a warm building. Similarly, other weather conditions have little effect on the set point. No experience has been obtained in a driving rain, but wet pavement, drizzle, and snow have little or no effect on the tuning point.

The hook of the handle was deliberately designed to allow the cane to be carried over the arm or hung up when not in use. Rugged materials were used in the construction of the cane, since a cane is subject to many kinds of accident. The shaft is particularly vulnerable. The electronic cane's hollow shaft, made of plastic-bonded Fiberglas, has proved to be almost indestructible, in addition to being light in weight.

Balance and weight are important to the comfort of the user. The present electronic cane weighs just 2 lb and was designed with the center of gravity below and approximately in the center of the hand. Thus no torque is required to hold it in position, and very little fatigue results.

TRAINING PROGRAM

It is easy to learn to use the cane. Both sighted and blind instructors have been used to teach blind travelers. In teaching over 50 trainees it has been found that well over 80 per cent can learn to travel safely with the cane in less than five ½-hr sessions. Less than 10 per cent have been completely dissatisfied with the electronic feature or have been unable to use it properly. About 2 weeks of practice with minimum supervision and a final checkout with an instructor constitute the training program. Many travelers have been able to use the cane successfully with only ½ hr of training and 1 hr of practice.

For the purpose of performing a large-scale evaluation program, 100 canes have been produced. This evaluation program will make possible the development of a complete training and practice procedure.

out, and tickling or even itching may be felt at others. Conversely, the sensations appropriate to any one sensory system can be produced by several distinct types of stimuli. For example, pressure on the outside corner of the eyelid will stimulate the visual receptors there and cause a spot to appear on the inside edge of the visual field.

Another proposition states that a given sensory system is capable of producing only a given category of sense qualities and cannot be replaced by another sense. If this proposition is correct, it is fruitless to try to replace vision by attempting to evoke visual sensations through stimulation of the skin.

Whether the different modalities of skin sensation are fully analyzed and differentiated at the periphery and transmitted along separate central pathways or whether the nature of the skin stimulus is transmitted in terms of a coded temporal pattern of impulses along a common central pathway has been the subject of recent research. Wall and Cronly-Dillon [28] have discovered a common-sense system of the latter type in the cat, and they describe some aspects of skin sensation in man that are most easily explained by assuming that part of our sensory pathway is made up of cells of the common-sense type. They suggest that systems of both types, as well as intermediate forms, may exist. Another example of a study of the interaction between stimulus parameters is the experiments by Bergstrom and Lindfors [5], which show how impulsive force and contact area are related in the perceptual manifold. The significance of these results in the presentation of information to the tactile sense is that, in order to code information along different stimulus dimensions, one needs to know the nature of the interactions between the sense modalities and the amount of jamming of one modality by another.

Also of interest is information on the sensitivity of the skin to pressure [14], vibration [13], and electricity [25], which can be found in the literature. While thermal and chemical stimulation is likely to be too slow for most communication applications and little is known in a quantitative way about textural variations as a possible dimension to be coded, pressure, vibration, air-jet, and electrical stimulators have been tried in communication systems.

A useful guide to the sensitivity of the skin at various locations is the two-point limen. This measure is the distance apart at which two points, simultaneously applied, can be felt. According to Vierordt's law of mobility,[3] this threshold is the sum of a constant, which is less the more distal the member, and a variable, which decreases for more distal positions within the member. Thus sensitivity increases with mobility—continuously toward the more mobile end of a member with an abrupt

[3] Boring [6, p. 477].

change in slope at a joint. It is interesting also to note that the spatial threshold for the direction of movement of a light pressure stimulus along the skin is only about one-fourth the size of the two-point limen. Moreover, whether movement occurred or not can be perceived for distances too small for correct perception of the direction.

Another concept of interest to the designer of a tactile communication system is the neural unit proposed by von Békésy [4]. This unit expresses as a function of position the amount of sensation evoked by a point stimulus. Essentially, this function consists of a central area of sensation surrounded by a refractory band in which sensation is inhibited. By superposition of this function, the sensation due to any stimulus distribution can be predicted. However, this is a static result, and one needs to consider the dynamics of the sensation and the stimulus for a complete description. Unfortunately, while it is known that the cutaneous senses adapt rapidly, the relations between adaptation, fatigue, and neural dynamics are not clearly understood.

THE KINESTHETIC SENSE

Bastian in 1880 defined kinesthesis as sensation arising from the tendons, joints, skin, and muscles [3]. This sense involves the recognition of position, of active and passive movement, and of resistance to movement. The sensory receptors responsible for these proprioceptive sensations are somewhat better understood than the receptors for tactile sensations. Four sets of receptors are involved, two in the muscle, one in the tendon, and one in the fascia associated with muscle. The "flower-spray endings," which are located in the muscle, are found to be stimulated by passive stretch of the muscle. It has been postulated [20] that the annulospiral endings located in the muscle spindles are sensitive to a difference between the extension of the main muscle and the muscle spindles. Both types of these receptors are involved in registering movement. The tendon organs of Golgi are in series with the muscle and respond to tension. Thus these organs are involved in the sensation of resistance to movement. The joint receptors are usually credited with the measurement of position and passive movement. (Unlike the cutaneous receptors, the kinesthetic organs mentioned so far adapt very slowly.) Finally, the Pacinian receptors in the fascial tissue probably report mechanical deformation caused by deep pressure or muscle movement, and the free nerve endings in the muscle, tendon, fascia, ligaments, and joints probably register deep pain.

The position sense is particularly important for transmission of information to the blind, for evaluations of space by the sense of touch are

out, and tickling or even itching may be felt at others. Conversely, the sensations appropriate to any one sensory system can be produced by several distinct types of stimuli. For example, pressure on the outside corner of the eyelid will stimulate the visual receptors there and cause a spot to appear on the inside edge of the visual field.

Another proposition states that a given sensory system is capable of producing only a given category of sense qualities and cannot be replaced by another sense. If this proposition is correct, it is fruitless to try to replace vision by attempting to evoke visual sensations through stimulation of the skin.

Whether the different modalities of skin sensation are fully analyzed and differentiated at the periphery and transmitted along separate central pathways or whether the nature of the skin stimulus is transmitted in terms of a coded temporal pattern of impulses along a common central pathway has been the subject of recent research. Wall and Cronly-Dillon [28] have discovered a common-sense system of the latter type in the cat, and they describe some aspects of skin sensation in man that are most easily explained by assuming that part of our sensory pathway is made up of cells of the common-sense type. They suggest that systems of both types, as well as intermediate forms, may exist. Another example of a study of the interaction between stimulus parameters is the experiments by Bergstrom and Lindfors [5], which show how impulsive force and contact area are related in the perceptual manifold. The significance of these results in the presentation of information to the tactile sense is that, in order to code information along different stimulus dimensions, one needs to know the nature of the interactions between the sense modalities and the amount of jamming of one modality by another.

Also of interest is information on the sensitivity of the skin to pressure [14], vibration [13], and electricity [25], which can be found in the literature. While thermal and chemical stimulation is likely to be too slow for most communication applications and little is known in a quantitative way about textural variations as a possible dimension to be coded, pressure, vibration, air-jet, and electrical stimulators have been tried in communication systems.

A useful guide to the sensitivity of the skin at various locations is the two-point limen. This measure is the distance apart at which two points, simultaneously applied, can be felt. According to Vierordt's law of mobility,[3] this threshold is the sum of a constant, which is less the more distal the member, and a variable, which decreases for more distal positions within the member. Thus sensitivity increases with mobility—continuously toward the more mobile end of a member with an abrupt

[3] Boring [6, p. 477].

change in slope at a joint. It is interesting also to note that the spatial threshold for the direction of movement of a light pressure stimulus along the skin is only about one-fourth the size of the two-point limen. Moreover, whether movement occurred or not can be perceived for distances too small for correct perception of the direction.

Another concept of interest to the designer of a tactile communication system is the neural unit proposed by von Békésy [4]. This unit expresses as a function of position the amount of sensation evoked by a point stimulus. Essentially, this function consists of a central area of sensation surrounded by a refractory band in which sensation is inhibited. By superposition of this function, the sensation due to any stimulus distribution can be predicted. However, this is a static result, and one needs to consider the dynamics of the sensation and the stimulus for a complete description. Unfortunately, while it is known that the cutaneous senses adapt rapidly, the relations between adaptation, fatigue, and neural dynamics are not clearly understood.

THE KINESTHETIC SENSE

Bastian in 1880 defined kinesthesis as sensation arising from the tendons, joints, skin, and muscles [3]. This sense involves the recognition of position, of active and passive movement, and of resistance to movement. The sensory receptors responsible for these proprioceptive sensations are somewhat better understood than the receptors for tactile sensations. Four sets of receptors are involved, two in the muscle, one in the tendon, and one in the fascia associated with muscle. The "flower-spray endings," which are located in the muscle, are found to be stimulated by passive stretch of the muscle. It has been postulated [20] that the annulospiral endings located in the muscle spindles are sensitive to a difference between the extension of the main muscle and the muscle spindles. Both types of these receptors are involved in registering movement. The tendon organs of Golgi are in series with the muscle and respond to tension. Thus these organs are involved in the sensation of resistance to movement. The joint receptors are usually credited with the measurement of position and passive movement. (Unlike the cutaneous receptors, the kinesthetic organs mentioned so far adapt very slowly.) Finally, the Pacinian receptors in the fascial tissue probably report mechanical deformation caused by deep pressure or muscle movement, and the free nerve endings in the muscle, tendon, fascia, ligaments, and joints probably register deep pain.

The position sense is particularly important for transmission of information to the blind, for evaluations of space by the sense of touch are

made in most cases by means of information supplied kinesthetically in the course of active exploration. The distance between two points is perceived by the amplitude of movement necessary in order to feel the two points. The sensitivity of this sense is emphasized by the fact that it is possible to detect, between the thumb and index finger, a difference of thickness of the order of 0.15 mm. Stevens and Stone [26] have determined scales for this type of perception.

Of particular importance from an informational-display standpoint are the joint receptors, which measure the angle of the limb, thus giving information about passive movement. Several investigators have measured the sensitivities of various body limbs to passive movement [7, 9, 17, 18, 23]. Thresholds for low-velocity movements range from 0.2 to 0.4° for the elbow and wrist to 1 to 2° for the ankle and great toe.

However, for information presentation through passive movement, there are several advantages for applying the stimuli to the hand. The hand is one of the most sensitive parts of the body, it has many degrees of freedom, and its size is such as to require equipment of moderate size and power. A basic finger movement, which might be used in a sensory-display device to convey discrete information, is a position pulse.

Some experiments have been performed by the author at the Massachusetts Institute of Technology to determine abilities of the human being to discriminate between position pulses of different amplitudes, durations, and directions. These experiments, exploratory in nature, used only a few subjects in a limited number of trials. The studies indicated the following:

1. A difference of less than 0.002 in. out of a total finger movement of 0.025 in., in which the duration of the pulse is 100 msec and the rise time is 15 msec, can be detected.

2. Varying pulse duration in the range of 55 to 135 msec seemed to have no systematic effect on the sensitivity of duration discrimination. However, finger position and direction of passive motion do seem to have an effect.

3. In the pulse-duration range of 55 to 135 msec and the pulse-height range of 0.024 to 0.036 in., a change in the area or energy of the pulse can be detected, whereas changes in pulse height cannot be distinguished from changes in pulse duration. Thus, a person can detect a change but cannot be sure whether the change is in pulse height or in pulse duration.

INFORMATION DISPLAYS

Somesthetic communciation systems have been built and tested by numerous investigators. These systems may be divided into systems that

excite primarily the tactile sense and systems that depend mainly on the kinesthetic sense. This section briefly reviews a few of these displays, with specific reference to information displays for the blind.

The tactile displays can be subdivided into those which depend on simple contact or pressure, those which use mechanical vibration, those which use electrical stimulation, and those which stimulate by use of an air jet. Braille is an example of a communication medium that depends on contact or pressure. Several machines have been built that automatically transmit braille, without the necessity of printing braille [8, 27]. In the reader designed by Bryce and Wheeler, tape punched in braille code passes over a sensing mechanism, which controls an electromechanical device that in turn sets up the braille by means of movable pins set in an endless plastic tape. As the belt passes across the reading station (an area approximately 10 in. long and 1 in. wide), the pins remain in their set positions, reproducing in braille the signs punched in the tape. The device built by Troxel differs from the Bryce-Wheeler braille reader in that the user plays a passive instead of an active part. In Troxel's device the perforated paper tape is used to valve air pressure that moves small poke probes against the fingers. There is a much greater separation between these probes than between the braille dots in a normal braille cell. Also, the fingers are placed in fixed position in this reader. Another type of communication depending on simple contact or pressure is the reading of embossed letters. Austin and Sleight [1, 2] have investigated discriminability of embossed letters by the sense of touch.

Various methods of using vibration as a stimulus for carrying information have been devised. Gault [11] attempted to apply speech energy directly to the skin, hoping that the similarity between the sense of touch and the sense of hearing would allow information to be transmitted through the skin in this way. While some limited success was achieved with the later versions of Gault's Teletactor, no communication system ever became generally practical. Part of the trouble is that it is difficult to distinguish between a change in intensity and a change in frequency or vibration. Recently Newman [21] has renewed this effort, by shifting the speech frequencies to a lower range that is more suitable for skin sensitivity. In addition, Guelke and Huyssen [15] used many small vibrators in such a way that frequency of speech was mapped into spatial location. Essentially a speech spectrogram was displayed across the fingers in the form of vibratory stimuli. Also, Geldard [12] has devised a system in which information is coded into intensity of vibration, duration of vibration, and bodily locus. This device consists of five buzzers located on the chest. A rate of about 38 wpm was achieved with this device.

Hawkes [16] has experimented with a cutaneous communication system based on the identification of different levels of electrical inten-

sity. Because of the small dynamic range between the threshold of feeling and the threshold of pain and because of the variability of skin resistance, avoiding pain is a problem with electrical stimulation.

The possibility of using small air jets as stimulation for tactile communications is being experimented with by the author and others at the Massachusetts Institute of Technology. One possibility along this line is an air jet that does handwriting on the skin. In one design, the author has rigged an *xy* recorder so that the pen mechanism carries an air jet of approximately 0.020 in. diameter. By application of the proper x and y voltages as a function of time, the recorder can be made to transmit handwritten information to the skin.

Information presentation to the kinesthetic sense is also being experimented with at M.I.T. Based on the experiments described in this paper, an eight-key information display for the fingers (excluding thumbs) has been designed and built, as shown in Figure 24-1. Each

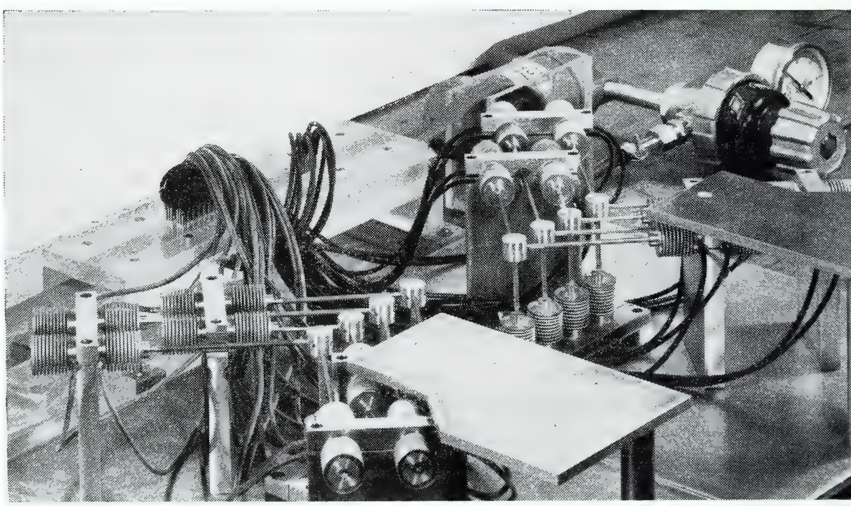


FIG. 24-1 *An eight-key information display for the fingers (excluding thumbs). Each key can be moved by a set of three orthogonal bellows assemblies, and each bellows assembly can either push or pull. Perforated paper tape is used to valve air pressure to the bellows.*

key is capable of movements of about ¼ in. in the $\pm x$, $\pm y$, and $\pm z$ directions. The keys are moved by sylphon bellows, and the air pressure can be valved by perforated paper tape. A simple computer program converts text material typed on a conventional flexowriter into one of several possible display codes. For example, one possible code is to associate each alphanumeral (letter of the alphabet or number) with a particular sequence of two movements. These movements may be pre-

sented to the fifth finger of the left hand, then repeated on the fourth finger of the left hand, while the next character is presented simultaneously to the fifth finger. This pattern is repeated so that a three-dimensional traveling wave of finger movements, representing a sequence of alphanumerals, moves across the fingers of both hands. Thus the observer can concentrate on a particular place and also be able to sense the peripheral symbols behind and ahead of the area of concentration. In this way the spatial aspect of the kinesthetic sense is utilized. Many variations of this display are immediately obvious.

CONCLUSION

What, if anything, can we speculate as to the ultimate information rates obtainable with tactile and kinesthetic displays? Tasks in which these senses are involved in the transmission of information (the reverse of using these senses to receive information) have been studied by several researchers [10, 24]. For example, Quastler and Wulff have investigated the limits of information transmission in typewriting (random texts) and playing piano by sight (random music). They obtained rates of about 15 bits/sec for typewriting and about 22 bits/sec for piano playing. Thus it appears that the motor-control system is capable of handling information at a rate not too much less than what we receive by way of our visual and auditory channels.

Once the optimum methods of stimulating the somesthetic senses for information transmission have been discovered, the problem of how to code the desired message into these stimuli remains. It will probably be necessary to develop more complex information-processing and coding methods in order to transform the message so that it is better matched to the channel and the human perceptual organization abilities. Of interest in this connection are the recent findings by Lettvin et al. [19] on the frog's eye. This research indicates that considerable organization and interpretation of information in the visual field are done in the retina. That is, four types of fibers were found in the frog's eye, each concerned with a different sort of pattern. Thus the image is expressed in terms of (1) local sharp edges and contrast, (2) the curvature of edges of dark objects, (3) the movement of edges, and (4) the local dimmings produced by movement or rapid general darkening. This is what is transmitted to the brain, not a more or less accurate copy of the distribution of light on the receptors.

Findings of this sort suggest that a promising approach to sending pictorial information over the somesthetic senses, as a substitution for vision, might be to have a series of filters which abstract pertinent infor-

mation from the visual field. Thus, instead of displaying the light intensity at each point in the field, considerable organization would be performed before the information would be displayed to the tactile and kinesthetic senses.

BIBLIOGRAPHY

1. Austin, T. R., and R. B. Sleight: Accuracy of Tactual Discrimination of Letters, *J. Exptl. Psychol.*, vol. 43, pp. 239–247, 1952.

2. ——— and ———: Factors Related to Speed and Accuracy of Tactual Discrimination, *J. Exptl. Psychol.*, vol. 44, pp. 283–287, 1952.

3. Bastian, H. C.: "The Brain as an Organ of the Mind," pp. 540–544, Appleton-Century-Crofts, Inc., New York, 1896.

4. Békésy, G. von: Neural Inhibitory Units of the Eye and Skin. Quantitative Description of Contrast Phenomena, *J. Opt. Soc. Am.*, vol. 50, pp. 1060–1070, 1960.

5. Bergstrom, R. M., and K. O. Lindfors: Experimental Demonstration of the Euclidean-Pythagorean Structure and the Quadratic Metrics in the Perceptual Manifold of the Cutaneous Tactile Sense, *Acta Physiol. Scand.*, vol. 44, pp. 170–183, 1958.

6. Boring, E. G.: "Sensation and Perception in the History of Experimental Psychology," Appleton-Century-Crofts, Inc., New York, 1942.

7. Browne, K., J. Lee, and P. A. Ring: The Sensation of Passive Movement at the Metatarso-phalangeal Joint of the Great Toe in Man, *J. Physiol.*, vol. 126, pp. 448–458, 1954.

8. Bryce, J. W., and J. N. Wheeler: Reading Apparatus (IBM Braille Belt Reader), U.S. patent 2,521,338, Sept. 5, 1950.

9. Cleghorn, T. E., and H. D. Darcus: The Sensibility to Passive Movement of the Human Elbow Joint, *Quart. J. Exptl. Psychol.*, vol. 4, pp. 66–77, Spring, 1952.

10. Fitts, P. M.: The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement, *J. Exptl. Psychol.*, vol. 47, pp. 381–391, 1954.

11. Gault, R. H.: Recent Developments in Vibro-Tactile Research, *J. Franklin Inst.*, vol. 221, pp. 703–719, 1936.

12. Geldard, F. A.: Adventures in Tactile Literacy, *Am. Psychologist*, vol. 12, pp. 115–124, 1957.

13. ———: "The Human Senses," John Wiley & Sons, Inc., New York, 1953.

14. Grinley, G. G.: The Variation of Sensory Thresholds with the Rate of Application of the Stimulus. I. The Differential Threshold for Pressure, *Brit. J. Psychol.*, vol. 27, pp. 86–95, July, 1936.

15. Guelke, R. W., and R. M. J. Huyssen: Development of Apparatus for the Analysis of Sound by the Sense of Touch, *J. Acoust. Soc. Am.*, vol. 31, pp. 799–809, 1959.

16. Hawkes, G. R.: Cutaneous Communication: Absolute Identification of Electrical Intensity Level, *J. Psychol.*, vol. 49, pp. 203–212, 1960.

17. Laidlaw, R. W., and M. A. Hamilton: A Study of Thresholds in Apperception of Passive Movement among Normal Control Subjects, *Bull. Neurol. Inst. N.Y.*, vol. 6, pp. 268–273, 1937.

18. Lee, J., and P. A. Ring: The Effect of Local Anaesthesia on the Appreciation of Passive Movement of the Great Toe in Man, *J. Physiol.*, vol. 123, pp. 56–57, 1954.

19. Lettvin, J. Y., et al.: What the Frog's Eye Tells the Frog's Brain, *Proc. IRE*, vol. 47, pp. 1940–1952, 1959.

20. Merton, P. A.: Speculation on the Servo-control of Movement, in "The Spinal Cord," CIBA Foundation Symposium, pp. 247–260, Little, Brown & Company, Boston, 1953.

21. Newman, Richard: The Feasibility of Speech Transmission Using the Skin as a Sensor, presented at Air Research and Development Command Seventh Annual Science and Engineering Symposium, Nov. 29–30, 1960.

22. Pieron, H.: "The Sensations. Their Functions, Processes and Mechanisms," pp. 23–24, Yale University Press, New Haven, Conn., 1952.

23. Provins, K. A.: The Effect of Blocking the Nerves to the Joints and Skin of the Index Finger on the Appreciation of Movement at the Metacarpophalangeal Joint, *J. Physiol.*, vol. 143, pp. 55–67, 1958.

24. Quastler, H.: "Information Theory in Psychology," Free Press, Glencoe, Ill., 1955.

25. Stevens, S. S., A. S. Carton, and G. M. Shickman: A Scale of Apparent Intensity of Electrical Shock, *J. Exptl. Psychol.*, vol. 56, pp. 328–334, 1958.

26. ——— and Geraldine Stone: Finger Span: Ratio Scale, Category Scale, and JND Scale, *J. Exptl. Psychol.*, vol. 57, pp. 91–95, 1959.

27. Troxel, D. E.: Tactile Stimulation of the Fingers as an Information Input to the Blind, unpublished master's thesis, Massachusetts Institute of Technology, Cambridge, Mass., August, 1960.

28. Wall, P. D., and J. R. Cronly-Dillon: Pain, Itch, and Vibration, *Am. Med. Assoc. Arch. Neurol.*, vol. 2, pp. 365–375, 1960.

Chapter 25

BAT-TYPE SIGNALS AND SOME IMPLICATIONS

*Frederic A. Webster**

During the years following the definitive establishment by Griffin and Galambos [3 to 5, 9, 11, 12] of the ultrasonic nature of the bat's echo-locating pulses, most studies on the echo-location mechanisms of bats were descriptive and exploratory. They established the adequacy of the bat's auditory system for most day-to-day problems of orientation, avoidance, and insect pursuit, and they answered the first broad questions asked by curious scientists and laymen: What do the signal pulses look like? What frequencies do they incorporate? From how far are usable echoes returned? How easily are they jammed by noise within the same frequency bands?

By the mid-1950s, research was producing clear evidence of an echo-location system of great interest and potential significance [6 to 8, 10, 13, 15, 16, 18]. Highly sensitive to weak echoes, extraordinarily impervious to interference jamming, and capable of intercepting insect targets with uncanny speed and precision, the bat was found to possess an integrated set of techniques that was phenomenally successful despite seemingly prohibitive problems. Indeed the interceptions could be executed accurately amid complex configurations, such as the leaves of trees, and even in the presence of seriously interfering noise. Man-made systems using like methods (notably sonar and radar, with their associated ultraspeed computations) often seemed slow, clumsy, and ineffective by comparison. Moreover, man-made systems are, at best, millions of times larger than

* 62 Coolidge Ave., Cambridge, Mass.

This work was sponsored in part by M.I.T. Lincoln Laboratory, operated with support from the U.S. Army, Navy, and Air Force.

Many individuals contributed significantly to the planning and execution of the work behind the present discussion. Important contributors were N. I. Durlach, H. E. Edgerton, D. R. Griffin, A. P. Lagon, J. J. G. McCue, K. D. Roeder, and A. E. Treat. David Cahlander of Lincoln Laboratory and Massachusetts Institute of Technology was responsible for all the electronic and photographic equipment and collaborated in the experimental work.

the pencil-tip size of the bat's processing system [2, 19]. To achieve performance approximating the bat's capacities, man-made systems must often employ processing steps operating a thousand times faster than the millisecond range of neurological operations typical of the bat's nervous system [17]. Since the bat performs so astonishingly with these severe limitations of equipment, might not its techniques encompass secrets that we can ill afford to neglect?

Beyond military and theoretical considerations there are other areas in which a bat's echo-locating mechanisms might provide significant clues. The bat has a set of devices that substitutes effectively for the visual systems of most higher animals. It works so well that a bat can live for decades through its use. Surely, the development of devices for the blind could profit from a clearer grasp of how the bat's system works [9].

However, the first efforts utilizing information about the bat's echo-location system were disappointing. Attempts to employ bat-type sound pulses for detecting buoys at sea and to utilize such pulses in blind-guidance devices in order to aid in the recognition of spatial configurations, obstacles, step-downs, and surface textures were often of no avail, primarily because these applications bypassed crucial elements. The arbitrary selection of isolated properties fails to make use of many of the essential dimensions and variations of the bat's subtle system. In general, also, there are critical differences between the bat's problems and those which the experimenters were attempting to solve. Knowledge of the signals and the problems alone is not sufficient, for beyond these are the bat's uniquely adapted reception devices [9, 16] and a highly specialized set of information-processing techniques.

In order to understand what the bat does with its signals, we must become familiar with the total frame of reference within which the signals are used. Hence the first step is to look briefly at the kinds of things that a bat does with its echo-locating mechanisms and to consider the ways in which a bat might use specific components of its echo-locating apparatus or technique.

WHAT DO BATS ACCOMPLISH WITH THEIR ECHO-LOCATING SIGNALS?

To survive, a bat must accomplish certain things. It must reach areas that provide food, water, roosting places, and places to hibernate, and within these areas it must find its way around, recognize objects and configurations, avoid obstacles, and capture the flying prey on which it lives. A close inspection of the bat's activities suggests that its perceptual problems are dealt with in terms of four ranges, long, medium, short, and what may be called contact, or tactile, range.

A typical long-range problem is that of a bat's guiding its migrations over hundreds of miles, e.g., from a specific summer colony to a given winter cave. Unfortunately as yet almost nothing is known about how such migrations are guided, and hence little can be said about the role of the echo-location mechanisms. Medium-range problems are those of local day-to-day navigation, e.g., the location of hunting areas or regions that provide places to roost. In these operations echo location is undoubtedly supplemented by important clues obtained by passive listening, olfaction, detection of temperature shifts and air currents, and other aspects of perception. Here again, however, little is known of the way in which such navigation is actually accomplished. Short-range echo location refers to direct guidance by the use of returns from objects and surfaces that are of the order of 1 to 10 ft away, the distances at which the echo-locating pulses of most bats are chiefly effective. These have been studied in some detail. Contact range refers to situations in which, under many conditions, echoes probably come back too quickly for complete auditory appreciation. For an object 6 in. away, for example, an echo comes back in 1 msec, an interval that is probably too short for full auditory recovery even after a brief (e.g., $\frac{1}{3}$-msec) pulse. At such close range, moreover, a bat commonly reaches out for a target with its wing tip and makes physical contact [20], or it detects with tactile senses the air-cushion effect of a surface below. Though hard-and-fast boundaries cannot be drawn, significant differences do appear to exist in the mechanisms by which problems at these different ranges are handled.

Most short-range problems pertain to path evaluation, specific avoidance, or specific approach. A bat spends much of its active life in finding its way through the intricacies of vegetation, the passages of caves, and the varied structures of barns and attics. This calls for an appreciation of complex configurations and the determination of suitable flight paths with respect to these. Although such path evaluation may be thought of as the limiting case of an increasing density of obstacles, a bat's handling of general surrounding configurations is often functionally quite different from its manner of dealing with specific, isolated obstacles. Specific obstacles may be either stationary, such as tree limbs, or moving, as, for example, another bat. Frequently obstacles are brushed in passing, but such close misses appear to be the result of accurate evaluation rather than miscalculation. Indeed, while shaving past with its body, a bat will often neatly lift its wing around an obstacle, thus using the shortest path to the space beyond. Although a bat commonly slows down, or even hovers, in certain complex-shaped spaces, it often flies through extremely complex configurations of foliage with astonishing speed.

Of course, not all objects are to be avoided. There are many objects, configurations, and surfaces that are highly valuable to the bat. Insects

are the most common useful objects, and typically these are flying, although in some instances they are resting on a surface or have just landed there. Common configurations to be approached are niches of rock used for roosting and crevasses that may indicate the entrance to a cave or other enclosed space. Surfaces useful to a bat are the rough, vertical surfaces (e.g., trees and rocks) that provide a place to land and the smooth, horizontal surfaces that indicate water for drinking.

Just how bats locate water is not established. Possibly this is a quest that is sometimes aided by vision. Up to this point we have talked as though the insectivorous bats were totally blind, when in fact they are not. It is known for example, that they can be taught to seek food under a large white marker instead of a large black one [1]. On the other hand, various tests suggest that insectivorous bats have little pattern vision and cannot see motion. For example, if a morsel of food is moved silently around in front of a hungry *Myotis* bat, the bat normally gives no indication whatsoever of detecting either the food's position or its motion. Moreover, at close range it will not use echo location. Frequently, if a held insect gives a brief buzz and is quickly moved aside, the bat lunges for the place where the insect was but is unable to follow it.[1] Many experiments have confirmed that vision is completely irrelevant to most of a bat's target pursuit and obstacle avoidance [8, 15].

What about the detection of water? Vision obviously plays a minor role under usual laboratory conditions. Normally a thirsty bat will attempt to scoop water from any reasonably smooth, horizontal surface, irrespective of the general level of illumination and the relative lightness of the surface. One observer, however, does report bats' preference for a "bright, shiny" surface [27]. Under natural conditions it is possible that a light area surrounded by relative darkness or a light area in the lower half of the visual field may signify a water surface that is reflecting light from the sky. Indeed, bats emerge at dusk, when the sky is still light, and commonly seek water immediately. Moreover, water is one of a bat's

[1] A bat's capacity for detecting flying insects by the insects' flight sounds occasionally introduces some complication into studies of echo-locating proficiency for the capture of flying prey. In one study it was noted that in the absence of insect-masking noise one *Myotis* bat commonly hung silently on the wall and waited until it heard the faint buzz of a passing fly [15]. When the fly was within 1 ft or so, the bat suddenly raised its head and started emitting pulses. If the fly was not too close or flying too complex a path, the bat then took off for an interception by means of echo-locating pulses. When insect-masking noise filled the room, passive listening was useless and the catch scores of this particular bat dropped slightly. However, for the most part, this bat shifted its mode of detection to active echo location. Other bats in the study actually increased their catch scores slightly when the insect sounds were masked; thus it was clearly indicated that for these bats passive listening for insect sounds was normally nonessential either for detection or for catching. *Eptesicus* bats sometimes appear drawn to insect swarms by passive listening.

most crucial survival needs, and it seems reasonable that a bat's limited vision might have an important function in this search.

BY WHAT METHODS DO BATS PUT THEIR ECHO-LOCATING SIGNALS TO USE?

This chapter is concerned primarily with the use of the bat's auditory system for the active echo location of specific objects and surfaces and most particularly with a systematic analysis of the relations between the signals emitted by the bat and the problem being solved. Besides the direct experimental approach, to be described presently, the comparative approach is both suggestive and instructive. Different species of bats have different signal structures, and the differences can be related to variations in the echo-location problems. A word of caution is due, however, since the situation is deceptive. For one thing, a given set of signals may be used for several different purposes and in several different ways. From the study of one particular use, such as the recognition of surface texture or minute structure, one cannot validly conclude that the signal of a particular bat was shaped by this requirement. It seems likely that adaptations shaped by one aspect of a bat's existence are drafted to serve many other purposes, and, though adequate, some of these adaptations may fall far short of perfection.

Nevertheless, it is reasonable to expect that specific and interrelated adaptations do exist between signal structure and a bat's mode of behavior. Some examples will help to illustrate this point. During normal flight, certain slower-flying bats, such as *Myotis keenii septentrionalis* and *Plecotus rafinesquii*, emit pulses of a relatively short duration (roughly 2 to 3 msec), whereas the faster-flying *Lasiurus borealis* and *Eptesicus fuscus* emit cruising pulses lasting several times as long (about 8 to 10 msec). The former tend to select smaller targets; the latter choose larger and, sometimes, faster-flying targets. For the most part, the former appear to detect targets at a distance of only 2 or 3 ft; the latter detect at a distance of 5 to 10 ft. As a matter of fact, the cruising pulses of these fast bats, typically about 10 msec in duration, would produce overlap between the outgoing pulse and the returning echo for any object within 5 ft. The strong outgoing pulse or the mechanisms associated with its transmission would probably render inaudible the extremely faint echo of a small insect at such a distance. It has been suggested, however, that the relatively uniform decrease of frequency between initial and terminal portions of the pulse is well suited to the use of echo overlap for determining range [22]. Thus the end portion of the outgoing pulse is of relatively low frequency and would not easily mask

the initial high-frequency portion of a returning echo. Frequency separation or the rate of beats between pulse and echo could thus be used as a measure of distance to an object; however, it is presumed that such a technique would apply mostly to larger objects. Measurements have revealed that during the pursuit of insects the outgoing pulse is normally shortened so as to avoid overlap. It is thus evident that systematic relations exist between length of pulse, speed of flight, and distance of object, but many of the details remain to be established.

Other properties of the signal besides pulse length enter the picture. *Plecotus rafinesquii* emits signals of far lesser intensity than the signals of other bats in its region; it compensates for this partly by the possession of very large ears as receiving antennas. The frequency of its signals is also lower. What the advantage might be of such low-intensity low-frequency signals at first seems puzzling. A possible explanation, however, may lie in the ability of many moths, favorite targets of bats, to hear the pulses of an approaching bat and make sudden, successful evasions [23 to 26, 28]. The soft-signaled *Plecotus* has been seen to approach evasive moths (*Galeria*) from behind without setting off evasive tactics until the last instant, if at all. Such moths would not even fly in a room where *Myotis* bats, which emit signals of high intensity and frequency, were present. Thus certain signal properties of *Plecotus* may represent an adaptation specifically evolved to facilitate the catching of moths that possess ears. By the same token, the ears of moths may well have developed to permit escape from bats. Although this is suggestive of a fascinating interrelation of evolutionary components, we must not draw final conclusions at this point.

Actually *Plecotus* does not represent the echo-locating bat with the lowest-intensity pulses. Certain neotropical bats emit signals of even lower intensity and relatively brief pulse duration [14]. These bats spend much time flying through dense vegetation and live on objects such as fruit, flowers, and terrestrial animals, which are large and slow as compared with the targets of true insectivorous bats. Moreover, some of the signals are characterized by a complex and shifting harmonic structure, which may duplicate in part the functions of the striking frequency sweep observed in the small insectivorous bats of the temperate zones.

It is the signals of these so-called frequency-modulated (FM) bats that will primarily be discussed. Before consideration of these bats, however, an important type of non-frequency-modulated signal, namely, the long, constant-frequency pulse of the Old World horseshoe bats (*Rhinolophidae*), should be mentioned. Many of their pulses last $\frac{1}{20}$ sec (corresponding to an echo distance of 25 ft) and are thus still being emitted while the echoes of chief interest to the bat are coming back. This fact, together with the remarkable constancy of the emitted fre-

quency, suggests that Doppler effects may well be used within a single pulse. This appears to be contrary to what is possible with short FM pulses. Another interesting feature of the echo-locating process of *Rhinolophidae* is the use of rapid oscillations of the external ears. The precise reason for the oscillation has not yet been determined. Possibly the action produces some type of induced modulation that supplements the direct superposition of echo on the outgoing pulse. It may aid particularly in the sharpening of angle evaluation. Indeed the number of echo attributes that a bat must use simultaneously in order to respond effectively in the short interval between detection and action suggests the need for a rapid segregation of incoming information into different analytical channels. The ear-oscillation procedure may be part of such a mechanism. However, again it must be pointed out that any such ideas at this stage are speculative.

A review has been presented of some possible relations between signal structure and physical situation as seen in the adaptive relations between the bat's signals and the environment in which they are used. At the same time it has been indicated that such relations are often far from simple. The catching of tiny insects, for example, may often have been a requirement of the utmost importance. Only if the bats could locate and catch targets the size of mosquitoes and smaller could they be sure of enough to live on. A wavelength of 3 mm, corresponding to 120 kc, would give appreciable echoes only from an object 1½ mm in diameter or larger [21]. A carrier frequency of the order of 100 kc would thus be mandatory in certain critical situations. Though lower signal frequencies may have many significant assets (e.g., lesser attenuation and easier generation), the importance of small-target detection may have overridden these.

The number of data that must be processed per unit of time, the precision needed, and the complexity of application (in terms of order or accuracy of prediction, number of attributes evaluated, etc.) govern the required rate of information handling. For example, the high-speed interception of a maneuvering target involves far faster information handling than simple obstacle avoidance. Greater precision of aim is essential, the timing must be more exact, prediction of a curving trajectory may be needed, the target echo is often weak with respect to the echoes of surrounding obstacles, and evasion tactics may have to be evaluated. Avoiding a stationary obstacle in an otherwise clear field requires only the elimination of a certain flight zone. Had the signals been evolved solely for simple obstacle avoidance, they might be far simpler than those required for the accurate interception of moving targets or the evaluation of complex spatial configurations.

Other important factors may also have entered into the evolution of the echo-locating signals of bats. One such factor is echo interference that occurs, for example, when scores of bats fly simultaneously in a space the size of an ordinary room. Typically the bats avoid collisions and find their way around successfully. How do they distinguish between the echoes from their own signals and those of neighboring bats? An important element here may be the total echo pattern, with respect to both the bat's own signals and the signals of other bats. Moreover, even though much exploration is often necessary, bats seem to learn in detail the structure of familiar spaces and thus are able to reduce greatly the number of elements that they must heed.

A further property of the FM bat signals is their remarkable resistance to jamming by random noise within the frequency band of the signals [2, 9, 13, 16]. Moreover, their suitability for clutter reduction and moving-target indications is clear from the successful pursuits of small targets among finely structured trees and shrubs. A mosquito flying close to the needle tips of a pine or spruce would produce an echo that is small and confusing in relation to the surrounding echo structure; yet frequent captures of these and smaller insects among the needles of such trees have been observed.

The success with which small insectivorous bats execute interceptions amid many potentially conflicting echoes suggests, among other things, an accurate gating of indications during the early stages of processing. In other words, the echo-locating system must have the capacity to label a particular anticipated time interval and other associated properties and to switch potentially conflicting indications out of the processing channels. Furthermore, relatively simple combinations of crude indications must often provide a basis for the quick initiation of suitable action that is characteristic of the bat's responses.

The interception courses of some bats indicate the accurate use of anticipatory planning. Certain wild bats have been readily trained to intercept meal worms projected upward 10 or 12 ft by a catapult. Sometimes these targets fall to within a foot or two of the ground before the bat can reach them; yet in such cases the bat will characteristically come in close to the ground and at a right angle to the falling meal worm, accurately predicting the point at which the worm will fall at the instant of interception. Moreover, the intercept course provides a smooth path between obstacles, both before and after the catch, even when some of these obstacles may be in new positions at the time of the catch.

Although these observations do not reveal specifically what the bat detects in the echoes or how the indications are so quickly transformed into effective patterns of action, they do give useful clues as to certain

elements that may be important. They suggest leads for many of the experiments that must be performed if one hopes to uncover the secrets of the bat's phenomenal echo-locating capacities.

THE NATURE OF FM-TYPE BAT SIGNALS

The preceding discussion pointed to factors that may have shaped the signals of insectivorous bats. It is clear that the relations between the signal and the associated physical problem are often complex. Certainly a long series of experiments, which is beyond discussion in the present analysis, is required in order to clarify these relations. It is intended here to give only a few preliminary and tentative suggestions as to the kind of relations that may exist and thus to put into rough perspective some of the areas in which useful experimental findings may occur.

Certain attributes of the bat's signals go with particular physical situations. These signal attributes and the bat's actions shift as the problem it is handling evolves. The bat's echo-locating problem can be divided into three broad phases, search or cruise,[2] detection and approach,[3] and terminal phase.[4]

During the search or cruise phase, the wing beat of a *Myotis* bat is even and symmetrical, and its flight path is straight or smoothly curved. Normally its head is directed straight ahead, and its mouth is wide open. During the detection and approach phase, however, the wing beat becomes intermittently asymmetrical and often irregular, with wing excur-

[2] There is some evidence to suggest that search and cruise may not be the same. Search implies a preparedness for pursuit, whereas cruise suggests a mode of flight that is not intended chiefly for the moment-to-moment detection of flying prey. Bats are frequently preoccupied with matters other than hunting and pay no apparent heed to insects in their path. There may well be distinctive differences between pure search signals and pure cruise signals. On a priori grounds one might expect a pure cruise phase to have pulses that facilitate detection and evaluation of obstacles and configurations at maximum distance and a pure search phase to have pulses optimizing some balance between certainty and distance of detection of expected prey. The former pulses might be of longer duration, lower frequency, and slower repetition rate. Of course, the pulses of so-called "normal flight" might well be specialized along continua other than these. For present purposes, however, such pulses may certainly be included under one general heading.

[3] It is sometimes convenient to divide the detection and approach phase into three subphases, initial detection, evaluation, and preparatory action. In most cases, however, these functions overlap, and current measures of such divisions are extremely uncertain. Furthermore, the existing divisions are defined in somewhat different ways by different observers.

[4] Terminal phase typically refers to a catch or attempted catch, although corresponding sets of signals and reflexes appear to apply to landing maneuvers.

sion generally increased, and the flight path is suddenly deflected unless the target is straight ahead. The bat's head and ears are generally directed accurately toward the selected target. Toward the end of the approach phase there may be a seemingly anticipatory curving forward of the tail, which possibly acts as a brake. During the terminal phase, the wing beat is suspended, and one or both of the wings may be used in target retrieval. The head is moved toward the target, with the ears sometimes tipped forward; the mouth is open and aimed for the target. For the catch proper the tail membrane usually comes forward over the face; sometimes, however, the target is first swept by a wing into the tail pouch.

Such maneuvers can be seen in detailed photographs of catches of meal worms tossed up to the bats in flight. These photographs are of two kinds, multiple-exposure pictures and high-speed motion pictures. Figure 25-1 shows the bat's accurate approach to a rising meal worm,

FIG. 25-1 *Bat's approach to rising meal worm.*

which is typical of many of the good catches that have been observed. Figure 25-2 shows the usual action of the tail membrane during a "direct" membrane catch. Note that one end of the meal worm is in the bat's mouth as the bat comes out of the catch about $\frac{1}{10}$ sec after snapping tail

FIG. 25-2 *Action of tail membrane during direct membrane catch.*

membrane and head together for the catch proper. Another (single-exposure) view of a bat going into a catch is given in Figure 25-3. As has been mentioned, many catches are made by reaching out with a wing and scooping the target into the pouch of the tail membrane before seizure with the mouth. Many variations of catch technique are used to produce the bat's remarkably high catch scores (sometimes over 90 per cent).

Table 25-1 indicates the distinctive attributes of FM bat signals.

The attributes pertaining to the overall configuration of the pulses or their sequencing can be observed roughly in the rectified pulse indications recorded by sound-on-film cameras. These attributes include (1) the pulsed form, with interpulse quiet interval, (2) variable pulse duration, and (3) variable pulse-repetition rate. The attributes that pertain chiefly to the detailed pulse structure include (1) high-average carrier frequency, (2)

FIG. 25-3 *Bat catching meal worm.*

significant and rapid frequency sweep, from high to low, and (3) variable frequency span.

The signal system of a representative bat (*Myotis lucifugus*) appears to operate between two extremes, cruise and buzz. The cruise extreme seems to be designed to provide maximum essential information on distant objects by means of a single long pulse, whereas the buzz extreme seems designed to provide a high and sustained data rate on close-range objects by the use of a rapid and regular sequence of brief pulses. In other words, during interceptions and landing maneuvers a bat's echolocating mechanism tends to shift from a system that gives complex

TABLE 25-1 *Some Signal Properties of FM-type Bats**

Signal property	Possible purpose or use	Possible advantage or utilization method
Envelope properties:		
Discrete pulsed form (with interpulse quiet)	Range determination	Measurement of time from pulse emission to echo return
	Clutter reduction	Separation of echo groupings
	Increased effective processing speed	Simplification of required steps and equipment
Variable pulse duration	Increased real-time adjustability of signal to external problem	Long pulses: *a.* Higher energy output per pulse (with resultant stronger echo) *b.* Increased bandwidth (assuming limited sweep rate), hence more precise time reference for echo† *c.* Possible Doppler use in some bats where steep initial frequency drop is combined with long, flat terminal segment Short pulses: *a.* Prevention of undesired transmit-receive overlap *b.* Increased clutter separation *c.* Faster range-data availability
Variable pulse-repetition rate	Increased real-time adjustability of signal (Variabilities in duration and rate probably aid in balancing the requirements for speed, precision, and processing simplicity)	Low repetition rate: suppression of ambiguity between returns from different ranges High repetition rate: *a.* Increased data rate *b.* Greater average energy at reduced pulse length
	Increased resistance to interference	Variable grouping of pulses. Possible limited pulse code modulation
Carrier properties:		
High average frequency	Detection of small objects	Rayleigh scattering from small objects increases with frequency
	Evaluation of surface structure	Back scattering from rough surfaces increases with frequency; also effects from detailed structure are sharpened
	Improved determination of relative elevation and azimuth	Beam width of transmit-receive system narrows with increasing frequency

TABLE 25-1 *Some Signal Properties of FM-type Bats* (Continued)*

Signal property	Possible purpose or use	Possible advantage or utilization method
Carrier properties (cont.): High average frequency (cont.)		Head shadows give sharper left-right intensity differences at higher frequencies (which in turn may amplify effects of time difference of arrival at the two ears)
		Effects due to form, position, and action of the outer ears increase with frequency
	Improved detection of signals in the presence of noise	Directional sharpening with increasing frequency permits better focus on a specific echo region
	Possibility of Doppler use for relative velocity	Frequency shift between signal and echo increases with frequency
Large frequency sweep:		
Effects due to large bandwidth	Improved accuracy and resolution in range determination	Time reference is sharpened with increasing bandwidth† (hence improved initial evaluation of range and improved time gating for selection among multiple ranges)
	Improved accuracy and resolution in azimuth determination	Interaural time differences probably reach usable degree of resolution (bat's small interaural distance limits possible use of time differences)
	Improved resistance to noise interference	Signal energy spread over wide band is resistant to masking by narrower bands of noise; also, matched filter would not act to shift noise components to a uniform phase reference‡
	Improved evaluation of surface structure and angle	Suppression of ambiguities due to unspecified phase reference; also, echo strength with shift of incident angle varies with frequency
Effects due to frequency sweep proper	Simplification of processing for size, texture, and structure	Since reflected properties vary with signal frequency, different indications from initial vs. terminal portions of pulse can be interpreted as due to the frequency differences

TABLE 25-1 *Some Signal Properties of FM-type Bats* (Continued)*

Signal property	Possible purpose or use	Possible advantage or utilization method
Carrier properties (cont.): Effects due to frequency sweep proper (cont.)	Possible facilitation of reception during transmission	Under conditions of signal-echo overlap, high-frequency portion of echo would overlap lower-frequency terminal emission, with reduced masking effects
	Possible supplementation of data on relative angle	Transmit-receive gain patterns vary with frequency and may give characteristic amplitude pattern, during frequency sweep, as a function of relative angle
	Possible improvement of signal detection in noise	Assigned frequency sweep may give identifying tag to received sound patterns of primary interest
Variable frequency span	Adjustment of average beamwidth for focusing or defocusing search zone	High frequencies give sharper beaming and shadow effects to improve long-range accuracy at low data rate; low frequencies facilitate wide-angle reception during close range maneuvering
	Reduction of ambiguity in identification of echoes from different ranges	Progressive shift in emitted frequency range may provide reference scale for successive echoes

* This table represents only a most preliminary and hypothetical outline of some of the advantages that a signal system of the FM bat type might make possible. It is possible that the system is governed by a few predominant features and that many of the theoretical attributes are impractical or unrealizable. Thus the terminal buzz may be basically the only transmitting technique, compatible with the bat's inherent anatomy and physiology, that would provide the requisite high data rate.

† Statements indicating a sharpening of time reference with increasing bandwidth presuppose that the bat's processing system capitalizes on the bandwidth provided by the frequency sweep—in particular, that the auditory system can act as a matched filter, namely, a filter that acts to peak all frequencies at one phase reference.

‡ If noise is wide-band, the deterioration due to the required wide-band receiver is compensated for by the fact that the matched filter acts selectively, in effect, to move frequency components of signal echo to a uniform phase reference.

evaluative data on distant objects by means of a single pulse to a system that uses the pattern of returns from a train of simple pulses to control quick, reflex-type actions at close range. Possibly the processing during the search mode operates somewhat in the manner of a wide-angle gunsight with telephoto center, whereas the processing associated with

the buzz mode operates in the manner of a simple wide- or medium-angle lens.

Figure 25-4 is a somewhat idealized representation of the manner in which some of the attributes of a *Myotis* signal might shift during an

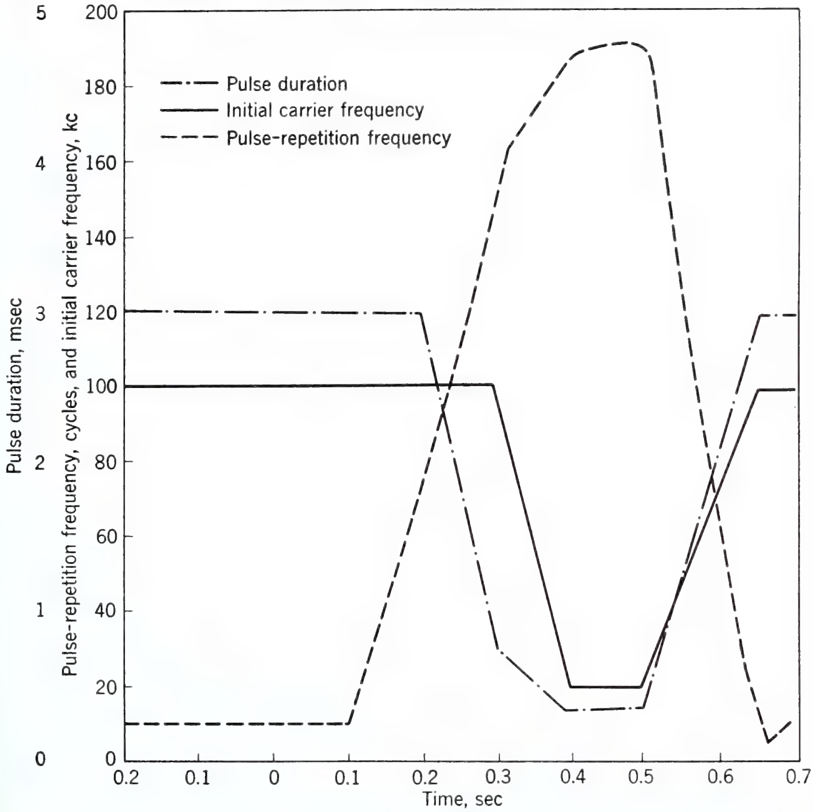


FIG. 25-4 *Idealized signal shifts of a Myotis lucifugus during interception.*

interception lasting ½ sec. Note that the climb in pulse-repetition rate is the first definite shift. Next comes a drop in pulse duration, and finally there is a drop in initial frequency. Since the frequency span per pulse commonly remains for a while of the order of one octave, whereas the pulse duration shortens markedly, the rate of frequency drop often increases significantly at this point. Later the frequency drop may lessen considerably. Enormous variations occur in certain aspects of the signal from interception to interception, as might be expected with the highly varied interceptions that these bats execute. The interception signals of other bats are often strikingly different. For instance, recordings of the red bat *Lasiurus borealis* show little, if any, decline in initial pulse fre-

quency from cruise to buzz. Possibly the much higher flight speed of these bats demands, at all stages of interception, the more precise information made possible by the echoes from sustained high-frequency signals.

HYPOTHETICAL ANALYSIS OF AN INTERCEPTION

Of course, any real understanding of the signals used by bats depends initially upon adequate experimental establishment of the purposes served by specific aspects of the signal. If these can be determined, then one can begin to construct a flow diagram that represents, to a useful approximation, the functional sequence of operations used by the bat in solving its echo-locating problems. In the present discussion only a first step toward such an approach is given. A typical echo-locating problem will be illustrated in some detail.

Figure 25-5 shows a complete interception of a fruit fly by a little brown bat, *Myotis lucifugus*. The diagram is a rather smoothed and simplified version of a successful interception described in an earlier

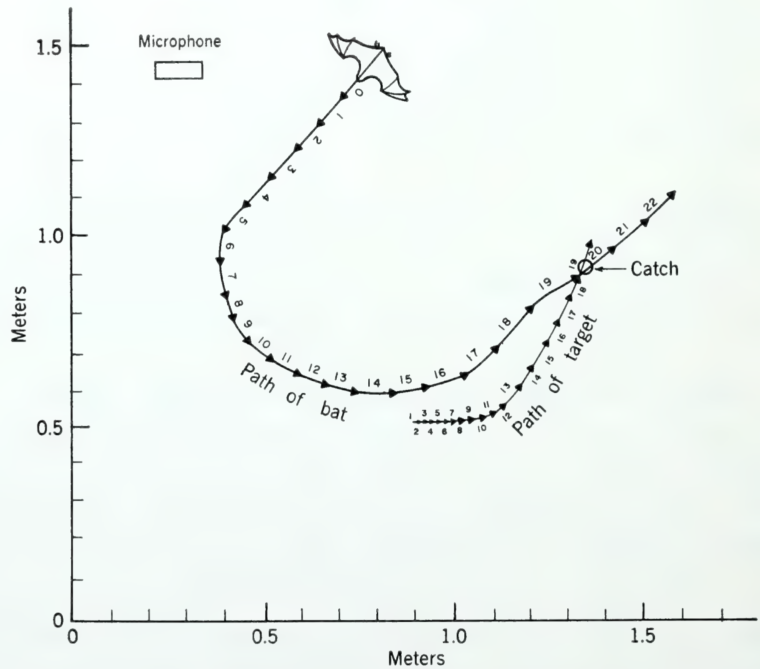


FIG. 25-5 *Path of bat, Myotis lucifugus, and target during interception. (Each vector arrow represents 1/24 sec.) (After D. R. Griffin, F. A. Webster, and C. R. Michael, The Echolocation of Flying Insects by Bats, J. Animal Behavior, vol. 8, 141–154, 1960. By permission of the publishers.)*

discussion [15]. The bat, traveling at about 2 m/sec (roughly 4 mph), is shown approaching the corner of a flight room in which several dozen fruit flies are flying. The path of the bat is represented by the sequence of arrows originating in the upper right quadrant and looping back toward that sector; the path of the target that is being intercepted is shown by the sequence of arrows just below. A potential specific obstacle, a recording microphone, is shown at the left. Frequently during these experiments many more obstacles, both stationary and moving, were present, and a large number of potential targets were within range. The situation illustrated, therefore, is far simpler than many interceptions executed with near-perfect success and at rates of up to two catches per second. The bat's capacity to deal quickly with complex situations is certainly far greater than the present analysis suggests.

Figure 25-5 was derived from data obtained through the use of two sound-on-film cameras, which were placed 1 m apart. Each arrow represents $\frac{1}{24}$ sec (roughly 42 msec), the frame interval of both the stereo films. Only a top view is given, since in this instance the interception was essentially level. At frame 6 the bat initiated a rather sudden left turn, of roughly 1 ft radius, that was continued until frame 16, about $\frac{1}{2}$ sec later, i.e., until just before the catch. At this point the turn was suddenly sharpened and the flight accelerated. The abrupt deflection shown between frames 19 and 20 is an example of the quick breaks in flight path that occur during catch maneuvers, in which sudden, violent, and asymmetrical action of the wings often occurs. Although the film is too blurred to permit judgment, the catch proper may have been executed with the left wing, since there is some indication that at the last instant the bat aimed too far to the right to allow a catch with the tail membrane, the more usual procedure. After the catch, the bat continued in a northeasterly direction.

The path flown by the target is indicated just below the path of the bat. Initially this fly moved slowly eastward; later its path deflected to a northeasterly heading and accelerated to 1 or $1\frac{1}{2}$ m/sec, a velocity that approached the speed of the bat. This increased target velocity may be of some interest here. It is possible that a bat normally treats a target of this size as stationary until near the catch and that even then it may assume some typical fixed velocity that gives an adequate approximation. Although the terminal action of the fly was unclear in the film, there is some evidence that the target overflew the trajectory expected by the bat at the time when its body aim was committed. The bat, however, has with the use of its wings a catch radius that extends several inches to each side of center. In addition, the bat's target-evaluating mechanisms apparently direct ear-and-mouth aim and control catch reflexes long after the aim of the body is essentially unchangeable.

It is suspected that at each stage a bat determines just enough information for initiation of effective procedures for the next stage. If some simple indication can satisfactorily replace some detailed analysis, then the simpler and quicker method is used. It seems that only by the simplest and most effective utilization of representative information is it possible for the bat to achieve such a high probability of catch in the quick interceptions that have been observed.

Figure 25-6*a* shows the interpulse interval as a function of time.[5] This record is typical of a direct approach, in which the bat proceeds to the point of intercept without relinquishing continuous echo-locating contact with the target.[6] Just where detection takes place is unclear. When a bat initiates a turn, as it did in the vicinity of frame 6, it commonly increases its pulse-repetition rate, as seen between frames 6 and 9. However, it is suspected that the sharp turn indicated here was initiated because of the bat's awareness of targets in the direction of the turn. Therefore, in this instance it seems likely that the bat detected weak echoes at frame 3 or 4 and that these echoes led the bat to the sharp deflection of its flight path at frame 6.

Figure 25-6*b* illustrates the actual distance between the bat and the target as a function of time and the corresponding time for the echo to return. It should be noted that after detection the bat continued to close the distance without significant variations.

To illustrate a systematic method of handling bat-interception data, Figures 25-7 to 25-13 present a hypothetical stage-by-stage analysis of the interception that was shown in Figure 25-5. The analytic intervals are each three film frames long (about 125 msec) and represent a rough approximation of what may be called a complete analytical response interval, i.e., the time required for a bat to process and respond to a characteristic set of incoming data.

Actually there is evidence that at least three response periods may exist. The longest, of the order of $\frac{1}{10}$ sec, appears to be concerned chiefly with the control of body aim and path. The immediate interval, of about $\frac{1}{15}$ to $\frac{1}{20}$ sec, seems to relate to control of the emitted signal and the receiving system (head and ears) and to preparation of the catch mechanisms. The shortest interval, perhaps of the order of $\frac{1}{30}$ sec, deals almost exclusively with catch reflexes. During this interval a bat

[5] Because of a slightly incomplete sound record in this particular catch, the initial points, indicated by open circles, have been taken from another case.

[6] There are times when a bat apparently detects a target too far to the side, too far above or below, or while traveling fast at too close range for a profitable direct attempt at interception. In such cases there is an abrupt rise in repetition rate while the bat directs its head toward the target. The rate then drops while the bat executes some maneuver to bring its body in line for a good interception run.

apparently combines the immediately preceding auditory indications with contact, or tactile, impressions. High-speed photographs of catches involving elements that were unexpected by the bat give some evidence that about $\frac{1}{30}$ sec may be required for the bat to appreciate wing contact. However, at this stage of research, all these times are largely guesswork.

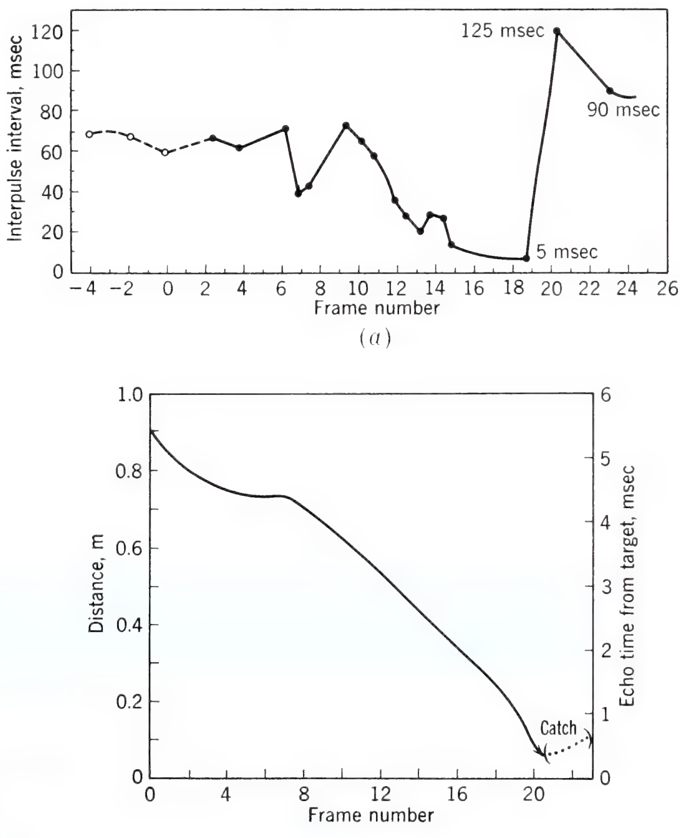


FIG. 25-6*a* *Interpulse interval as a function of time.*
FIG. 25-6*b* *Distance from target (with echo times).*

In addition, it should be noted that, although each interval may have a specific function, the processing of data observed in one interval continues during subsequent intervals and the accuracy of observation is thus increased.

Figure 25-7*a* is a positional diagram of the last segment of the cruise phase. The blips labeled *a* and *b* along the bat's flight path represent the points at which the bat emitted signal pulses. The accompanying time-

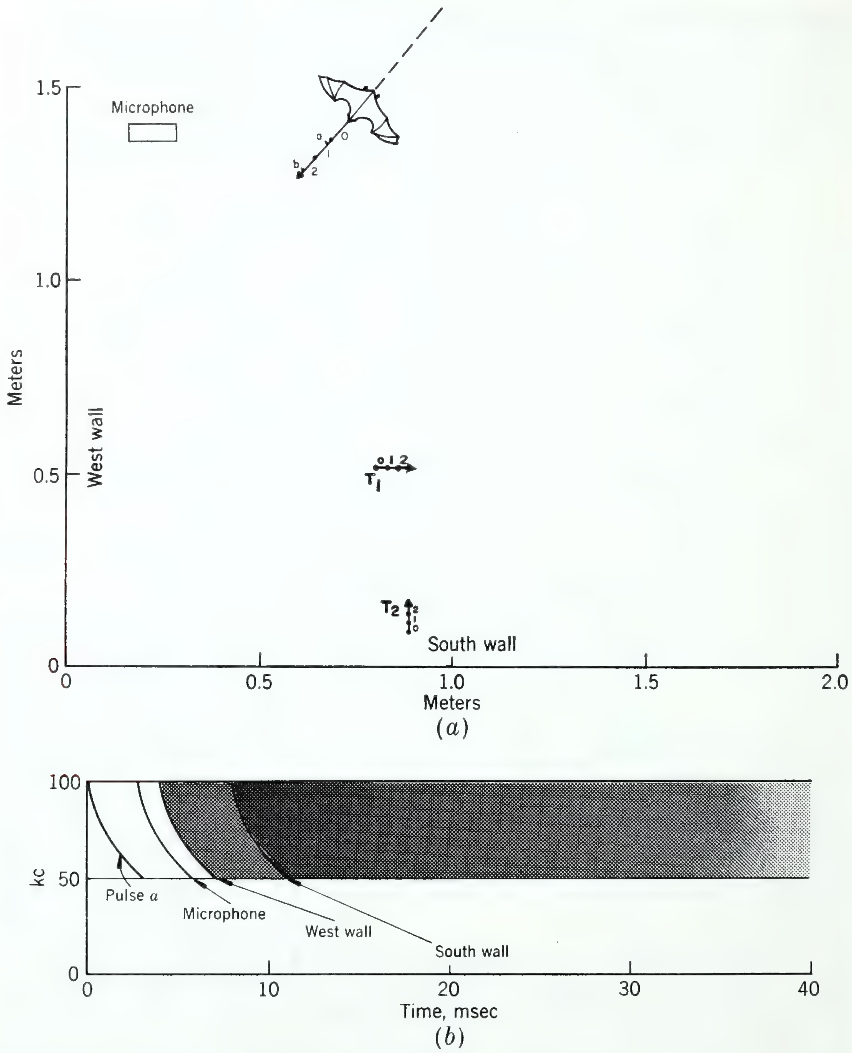
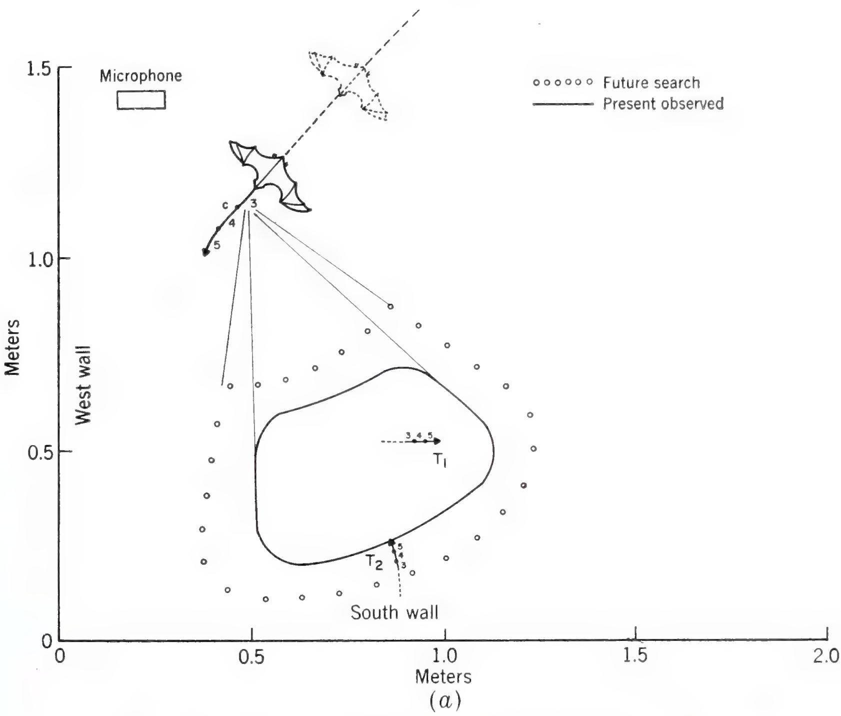


FIG. 25-7a *Positional diagram of interval 1, last segment of cruise phase.* **FIG. 25-7**b *Time-frequency-intensity plot.*

frequency-intensity plot (Figure 25-7b) shows the outgoing pulse a and its returning echoes as the bat might receive them. Since both potential targets T_1 and T_2 are too distant to provide detectable returns, no target echoes are included. Of course, the echoes from wall surfaces fade gradually from some maximum that is determined by distance, direction, orientation, and texture. The interval between cruising pulses is about 70 msec.



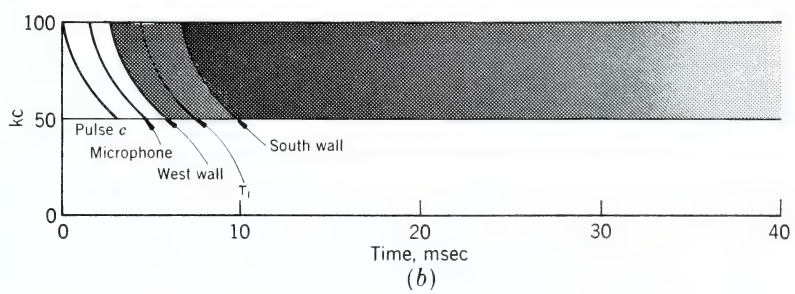


FIG. 25-8*a* *Positional diagram of interval 2, initial interval of detection phase.*
FIG. 25-8*b* *Time-frequency-intensity plot.*

Figure 25-8*a* is a positional diagram of interval 2, the initial interval of the detection phase. For the purpose of illustration, it is assumed that detection took place as the result of a detectable target echo from pulse *c*, which is shown in the accompanying time-frequency-intensity plot (Figure 25-8*b*). It is probable that the bat's initial observation is rather rough. However, just what the bat might discern through such a weak return from a single pulse, or even from two such pulses, is not established. Because the bat turns quickly in the correct direction, it is

assumed that it gains some idea of availability for catch and of the relative azimuth and elevation of the target, although perhaps only in such broad binary or ternary categories as left, as against ahead or right, level, as against up or down, far, as against near, small, as against medium or large, and unobstructed, as against blocked.[7] It is difficult to see how a bat could achieve as many as two precise catches within a second unless it could gain at least some such rough approximations from a single detecting pulse.

How might directional information be obtained from a single return? As yet it is known only that the structure of the signal provides for reception of such information. It is not known which of the possible signal attributes a bat actually uses or how it uses them. However, certain pertinent facts about the signal and the reception system have been suggested, as indicated in Table 25-1. First, the high frequencies of the first part of the pulse are likely to give sharp left-right amplitude differences. Second, the left-right ear separation of about 1 cm provides for interaural time differences of up to 30 or 40 μsec in arrival of echoes (and under optimum conditions a bat might discriminate down to 10 or 20). Third, the nature of reflection from the bat's outer ear varies at both ears with frequency. Fourth, the radiation pattern varies with frequency. Range data may be derived in several ways but perhaps chiefly through a measure of time-to-echo.

One way of thinking about a possible quick and simple processing method is to imagine that certain rough, rapid measures are made in binary form and then passed into a matrix system that gives fixed outputs according to the detected combinations. For a weak return the measures might take longer, since elements acting as filters may be required; however, by some such means certain appropriate categories of initial response might be simply derived. Thus reception at left and right ears might be compared, to two-value levels, with respect to average amplitude, upper-frequency amplitude and shift, lower-frequency amplitude and shift, and time difference of arrival. Certain overall measures, such as overall time-to-echo and echo discreteness, would also feed into the matrix system. The outputs of such a system would then switch into suitable sequences of motor response. Obviously, in presenting such a theoretical and unverified picture it is intended only to indicate that there are techniques capable of avoiding long processing sequences while at the same time providing an adequate basis for effective initial response.

Whatever the actual mechanisms involved, it seems likely that when it receives the echo from pulse *c*, the bat's functional concept of the

[7] It is possible that certain bats, such as the fast-flying *Lasiurus borealis,* also gain some idea of relative velocity, e.g., slowly approaching, as against rapidly approaching.

target and its location may be described in some such general terms as small, far, left, level, and unobstructed. The category small might also release responses associated with slow targets; likewise, the category unobstructed might release direct, rather than delayed, interception procedures. In Figure 25-8*a* the solid-line box containing T_1 indicates a slice through the approximate volume in which the bat might initially guess the target to be. This guess would provide the basis both for the initiation of pursuit and for the selection of the area in which the next observation should be made in order to gain further and more precise data. This next observation area, or search window, would typically be larger than the initial estimated target area and is indicated by the area defined by open circles.

Figures 25-9*a* and 25-9*b* represent interval 3, the initial interval of observable action by the bat. The actions and results observed on the sound film were wing action (right wing up, left wing down and in); head and ears turned in direction of target; signal shift (increased repetition rate), as seen previously in Figure 25-6*a*; body aim and flight path (body turning in left arc of about 1-ft radius). Presumably the bat has by now committed itself to the pursuit of the primary target T_1 and must therefore gate out indications from the second target T_2 now coming within range. Bats are often remarkably persistent in the pursuit of a chosen target even in the presence of several closer targets giving far more conspicuous echoes. Certain zones of acceptance must thus be set up in the processing system and only values within those zones used for control of interception behavior. In the case of conflict, as might be the case here, the values more central in the acceptance zones would normally be used. Once a zone of search has been established, it is possible that like processing methods are repeated, but to closer tolerances, within the smaller zones of selection in order to provide more accurate estimates of position and search window. Figure 25-9*a* shows the possible new areas of estimated position and search window, as well as the original estimate of position (now represented by a dashed line).

By interval 4, represented by Figures 25-10*a* and 25-10*b*, the bat may well have narrowed the positional determination sufficiently to derive an estimate of target path. This may be facilitated by a continuation of positional processing beyond what has been designated as the full-processing interval. It should be recalled that only a limited number of discrete sampling points are illustrated, whereas the bat probably uses a much denser sequence of overlapping samples. In any case, the bat is now in a position to make a more accurate estimate of the expected next position, which is beyond the zone of observed position. The expected next position is marked by solid circles, which are surrounded by the future search area, indicated by open circles.

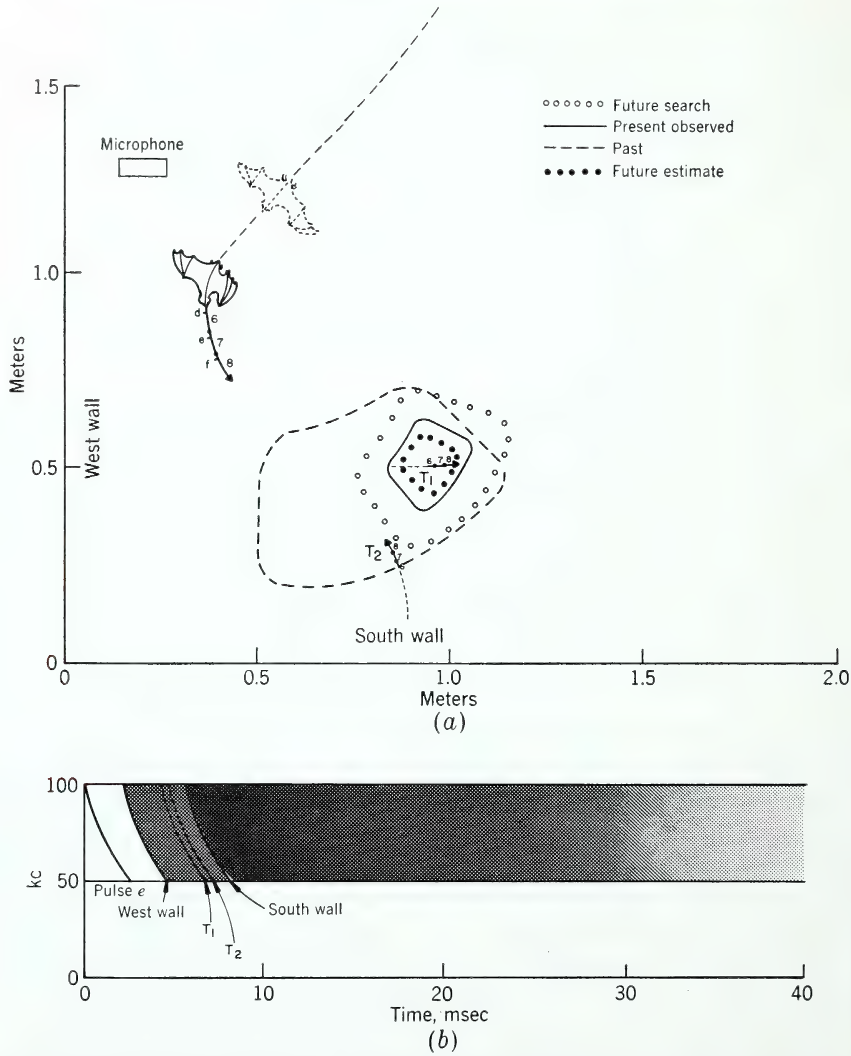


FIG. 25-9*a Positional diagram of interval 3, initial interval of observable action by bat.*

FIG. 25-9*b Time-frequency-intensity plot.*

At pulse *j*, shown in Figure 25-11*a*, the bat was probably unaware of the target's turn. The observed position (solid line) was thus to the left of the expected or estimated position (dotted line) and may have resulted in an appreciable sharpening of the turn that occurred between intervals 5 and 6 (Figures 25-11 and 25-12). In the time-frequency-intensity plot for interval 5 (Figure 25-11*b*) the shading is discontinued before pulse *k* in order to show the drop of successive pulse frequencies

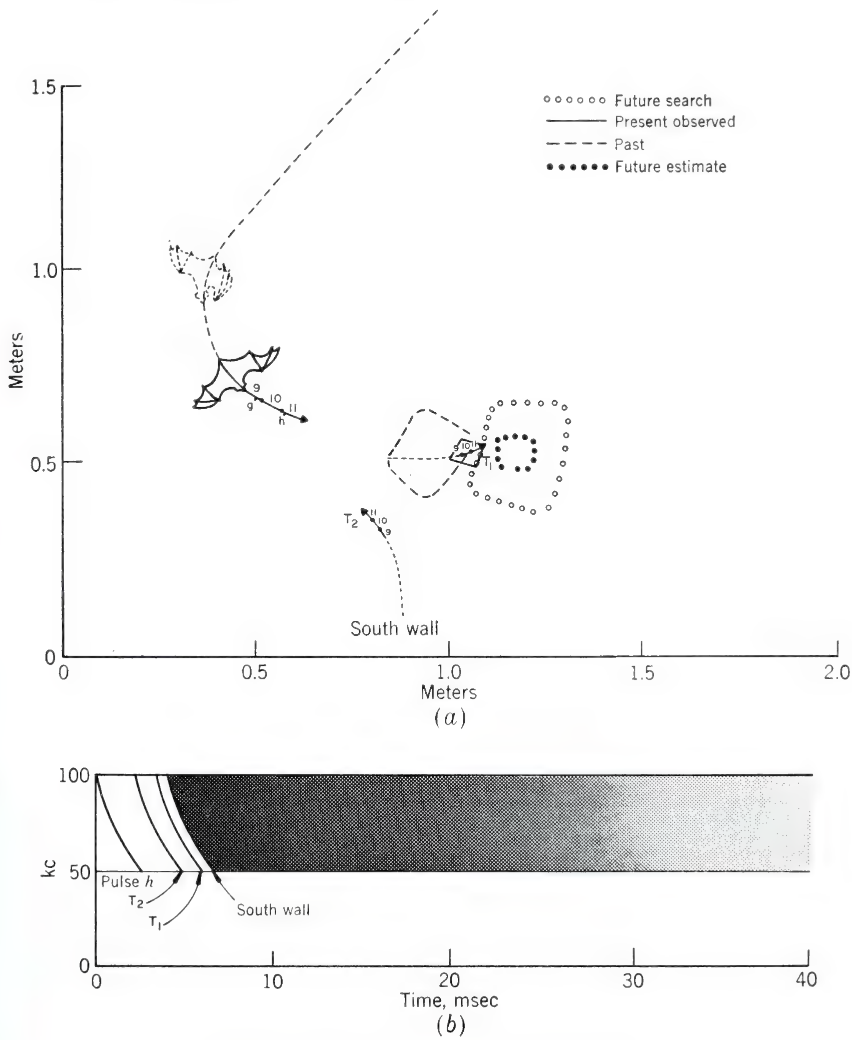


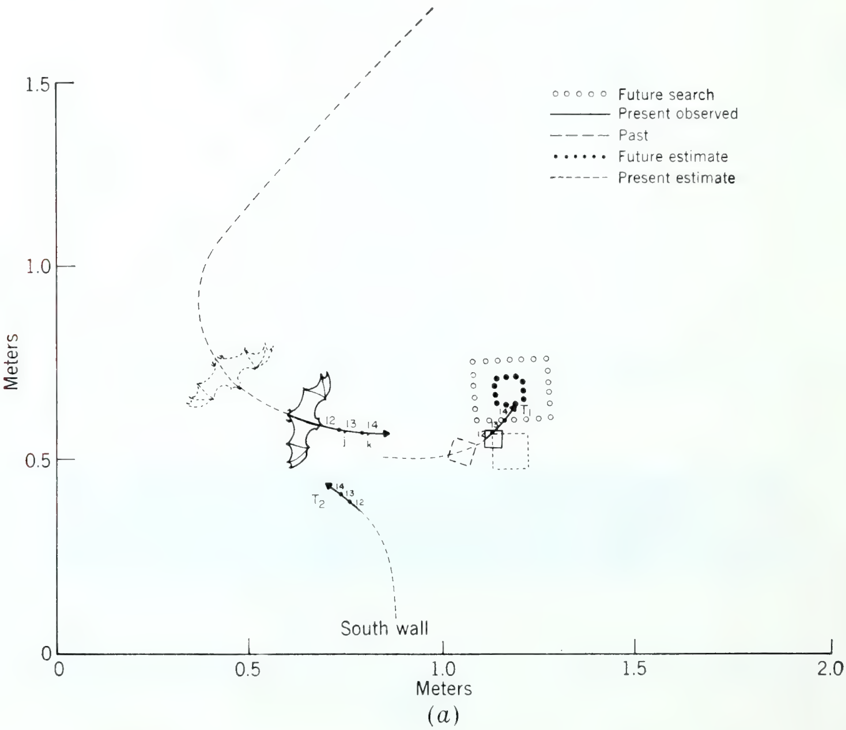
FIG. 25-10*a* *Positional diagram of interval 4.*
FIG. 25-10*b* *Time-frequency-intensity plot.*

into a zone below the frequencies of previous echoes. This drop may reduce echo ambiguity in closed spaces by placing the succession of echoes on an identifying frequency scale.

By interval 6 (Figures 25-12*a* and 25-12*b*) it seems unlikely that further indications could be used to modify significantly the planned path of the bat. However, by this stage the bat has undoubtedly derived a sufficiently precise idea of the target's position and path to be able to proceed directly to the anticipated point of intercept with virtual

certainty of catch. During interval 6 the signal underwent transition to the terminal buzz. A new kind of processing must take place here, and the results are probably used chiefly to govern the direction of the head and ears and to set in motion a suitable set of catch preparations.

Interval 7 and the catch proper are illustrated in Figures 25-13*a* and 25-13*b*. In the present catch the bat may have intended to catch the fly directly in the tail membrane, as is customary with small targets, but the fly appears to have attained a velocity greater than the bat's estimate.



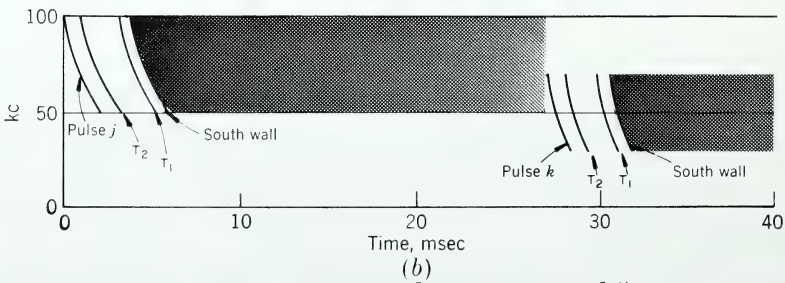


FIG. 25-11*a* *Positional diagram of interval 5.*
FIG. 25-11*b* *Time-frequency-intensity plot.*

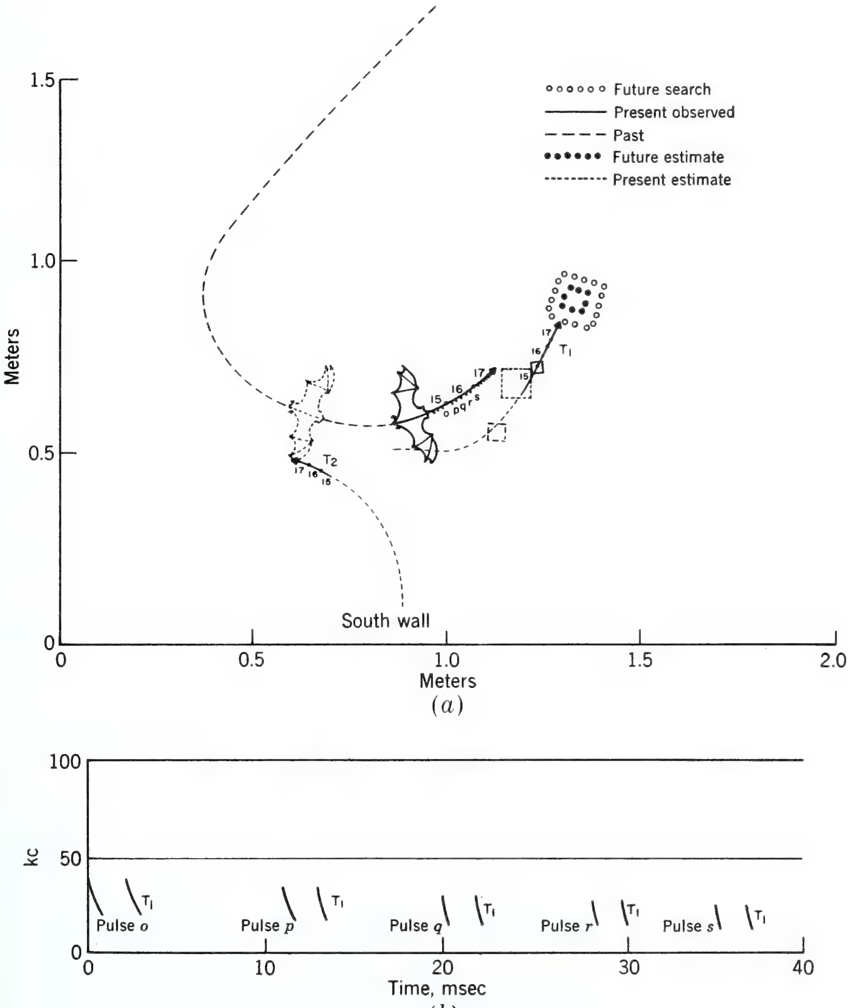
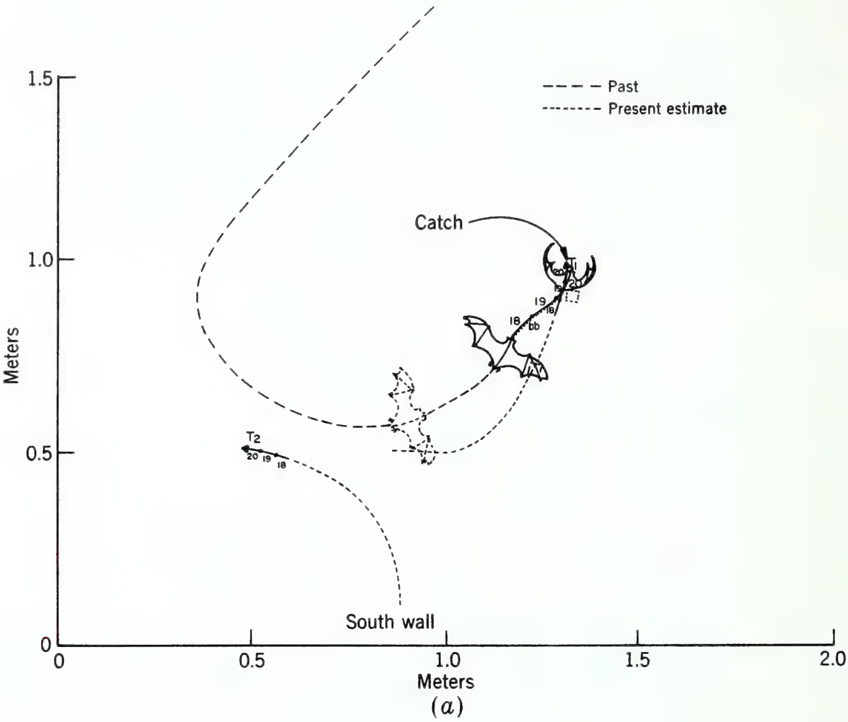


FIG. 25-12*a* *Positional diagram of interval 6.*
FIG. 25-12*b* *Time-frequency-intensity plot.*

Indeed, as suggested earlier, *Myotis* bats may simply assume a fixed average velocity for small targets or, in some cases, may treat the targets as stationary. Such a tactic would normally provide an adequate approximation. During the terminal buzz, however, the bat formed a more precise evaluation of position and realized at this juncture that a "wing-assisted catch" would be required. Certainly some sort of successful catch was made at frame 20.

In the present discussion, interceptions of airborne targets have been emphasized. Little has been said about obstacle avoidance or orienta-



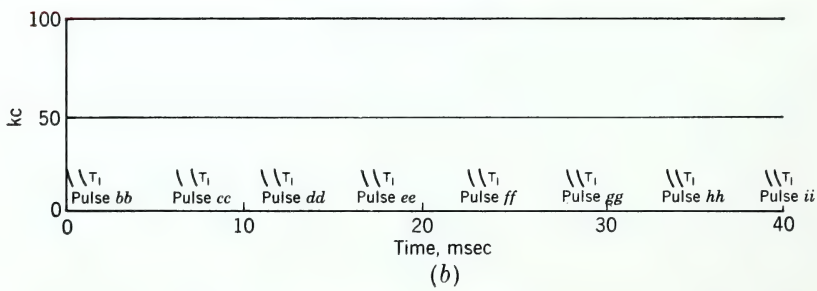


FIG. 25-13a *Positional diagram of interval 7 and catch.*
FIG. 25-13b *Time-frequency-intensity plot.*

tion within complex physical configurations such as the leaves and branches of trees. The uncanny speed and skill with which some bats fly through these intricate and unfamiliar spaces without collision and without confusion as to suitable path give evidence of a truly extraordinary capacity to evaluate the relevant gross properties of complex spaces. The determination of fine structure is evidently quite a different matter, for bats frequently make many passes by some configuration of interest before they decide how to deal with it. Clearly the bat's process-

ing system is designed for maximum speed where speed is imperative, other matters being dealt with by slower means.

In the illustration presented here there are obviously more uncertainties than certainties—uncertainties both as to the physical details of directly observable events and as to the true sequence of the processing details that guide the bat to its successful interception. Nevertheless, in a general sense, the sequence of events depicted in this chapter may well represent a first approximation of what occurs in such a catch. The information presented is, perhaps, the best guess that one can make at this stage. Of course, the entities considered, such as estimated position, observed position, and expected position, are hypothetical approximations of functional constructs that probably never exist in such pure form. Some such approximations, however, represent essential elements of a model that must be specified in order to achieve an effective experimental approach to the bat's actual methods of orientation and interception.

We have a long way to go before we shall be able to unravel the true secrets of the ways in which a bat deals with the intricate and bewildering indications that come back in a set of returning echoes. It is certain, however, that in the course of discovery much profitable information will be found.

BIBLIOGRAPHY

1. Eisentraut, M.: "Aus dem Leben der Fledermäuse und Flughunde," pp. 51–52, Gustav Fischer Verlagsbuchhandlung, Jena, Germany, 1957.

2. Freedman, J.: In Defense of the Bat, *Contemporary Psychol.*, vol. 4, no. 11, November, 1959.

3. Galambos, R.: The Avoidance of Obstacles by Flying Bats, *Isis*, vol. 34, pp. 132–140, 1942.

4. ——— and D. R. Griffin: Obstacle Avoidance by Flying Bats; the Cries of Bats, *J. Exptl. Zool.*, vol. 89, pp. 475–490, 1942.

5. ——— and ———: The Supersonic Cries of Bats, *Anat. Record*, vol. 78, p. 95, 1940.

6. Gould, E.: The Feeding Efficiency of Insectivorous Bats, *J. Mammalogy*, vol. 36, pp. 399–407, 1955.

7. ———: Further Studies on the Feeding Efficiency of Bats, *J. Mammalogy*, vol. 40, pp. 149–150, 1959.

8. Griffin, D. R.: Bat Sounds under Natural Conditions with Evidence for the Echolocation of Insect Prey, *J. Exptl. Zool.*, vol. 123, pp. 435–466, 1958.

9. ———: "Listening in the Dark," Yale University Press, New Haven, Conn., 1958.

10. ———: More about Bat Radar, *Sci. American*, vol. 199, no. 1, pp. 40–44, July, 1958.

11. ——— and R. Galambos: Obstacle Avoidance by Flying Bats, *Anat. Record*, vol. 78, p. 95, 1940.

12. ——— and ———: The Sensory Basis of Obstacle Avoidance in Flying Bats, *J. Exptl. Zool.*, vol. 86, pp. 481–506, 1941.

13. ——— and A. D. Grinnell: Ability of Bats to Discriminate Echoes from Louder Noise, *Science*, vol. 128, pp. 145–147, 1958.

14. ——— and A. Novick: Acoustic Orientation of Neotropical Bats, *J. Exptl. Zool.*, vol. 130, pp. 251–300, 1955.

15. ———, F. A. Webster, and C. R. Michael: The Echolocation of Flying Insects by Bats, *J. Animal Behavior*, vol. 8, pp. 141–154, 1960.

16. Grinnell, A. D., and D. R. Griffin: Neural Correlates of Bats' Resistance to Jamming, *Fed. Proc.*, vol. 20, p. 327, 1961.

17. ——— and ———: The Neurophysiology of Audition in Bats, *Anat. Record*, vol. 134, p. 574, 1959.

18. ——— and ———: The Sensitivity of Echolocation in Bats, *Biol. Bull.*, vol. 114, no. 1, pp. 10–22, February, 1958.

19. Licklider, J. C. R.: Nature as Sonar Engineer, *Contemporary Psychol.*, vol. 4, no. 1, pp. 1–4, January, 1959.

20. McCue, J. J. G.: Bats Hunt with Sound, *Natl. Geograph. Mag.*, vol. 119, pp. 570–578, 1961.

21. Novick, A.: Orientation in Paleotropical Bats, I. Microchiroptera, *J. Exptl. Zool.*, vol. 138, pp. 81–154, 1958.

22. Pye, D.: Echolocation of Bats, *Endeavor*, vol. 20, no. 78, pp. 101–111, April, 1961.

23. Roeder, K. D., and A. E. Treat: The Acoustic Detection of Bats by Moths, *Proc. Eleventh Intern. Entomol. Congr.*, 1960, in press.

24. ——— and ———: The Reception of Bat Cries by the Tympanic Organ of Noctuid Moths, in W. Rosenblith (ed.), "Sensory Communication," pp. 545–560, Technology Press, Cambridge, Mass., 1961.

25. ——— and ———: The Detection and Evasion of Bats by Moths, *Am. Scientist*, vol. 49, pp. 135–148, 1961.

26. ——— and ———: Ultrasonic Reception by the Tympanic Organ of Noctuid Moths, *J. Exptl. Zool.*, vol. 134, pp. 127–158, 1957.

27. Ryberg, O.: "Studies on Bats and Bat Parasites," Bokförlaget Svensk Natur, Stockholm, 1947.

28. Treat, A. E., and K. D. Roeder: Electrical Response of the Noctuid Tympanum to Ultrasonic Stimulation, *Proc. Tenth Intern. Entomol. Congr.*, vol. 2, pp. 117–120, 1958.

Part F

HUMAN FACTORS IN MANIPULATION AND MANUAL EXTENSION

Chapter 26

SOME CONSIDERATIONS IN THE DESIGN OF EXTERNALLY POWERED UPPER-EXTREMITY PROSTHESES

*Gerald Stone**

The engineering study of prosthetics encompasses numerous technical areas, wherein exists a variety of intriguing problems. Replacement of a body part involves the consideration of the physical device, communication, and motive power. In the normal individual there is an intricate, sensitive system that enables kinematic and kinetic function. In addition, there is a mode of control that provides for precise positioning of the extremity and for feedback of information about pressure, temperature, and position. The upper extremities, with their many degrees of freedom, constitute a system of such complexity that it is questionable that exact duplication of function and control by material means will ever be possible. It is believed, however, that by the application of sound engineering and medical techniques restoration of upper-extremity functions can be accomplished to provide for a much higher level of utilization than occurs in current practice.

In the present state of the art, upper-extremity prosthetics represents a severely limited degree of restoration of function. The prosthesis is largely dependent upon the remaining anatomy of the amputee, which is utilized for the control, force, and feedback functions. Therefore, the level of amputation is of prime importance in what can be accomplished by the prosthesis, since the higher the level of amputation, the greater the loss of function. For example, a long below-elbow amputation requires that the functions to be replaced are wrist flexion and hand prehension. An above-elbow amputation requires replacement of elbow stabilization, forearm flexion, forearm rotation, and wrist rotation, in addition to wrist

* Research Division, College of Engineering, New York University, New York, N.Y.

Appreciation is expressed to the Veterans Administration Prosthetics Center for the use of its extensive photographic collection, from which the photographs presented in this chapter were obtained.

flexion and hand prehension. However, even such functional kinematic restoration is not the only requirement of a prosthesis. Also necessary is replacement of the various communication channels and the basic muscular motive power that have been disrupted or removed.

In current practice, the simplest task appears to be the replacement of an upper-extremity limb with a device containing mechanical counterparts whose functions are similar to those of the normal anatomy. The problem of providing for the missing control-force-feedback relation is of another order. Difficulty in harnessing and coordinating enough separate bodily motions to enable operation of various prosthetic components presents complex problems that result in limitations of prosthetic function. The body manipulations employed for power and control usually lead to rapid fatigue, poor cosmesis, and related psychological and physiological difficulties. Feedback is mostly visual.

All these difficulties must be considered in any attempt to utilize sources external to the body to provide power and in any attempt to use normal body parts, not necessarily associated with the amputation site, for sophisticated control and feedback functions.

The ensuing discussion is intended to provide some insight into these problems associated with the construction of such externally powered upper-extremity prostheses. Separate consideration is given to each of the basic problems of kinematic- and kinetic-function restoration, control, and force and feedback restoration. Also considered are some of the methods of upper-extremity prosthetics that are the subject of current research.

BASIC BIOMECHANICAL CONSIDERATIONS

For the engineer to understand fully the problems involved in prosthetic restoration of the upper extremities, some knowledge of the basic body mechanism is required. By analogy with mechanical components the components of the upper-extremity mechanism are:

1. Bones—structural members
2. Joints—bearing surfaces
3. Joint linings—lubricants
4. Muscles—motors, damping devices, and locks
5. Tendons—transmission cables
6. Nerves—control and feedback circuits

These constituents are arranged in a manner that enables motion, with various degrees of freedom, and the execution of diverse activities, such as precise positioning in space, performance of external work, and fine digital manipulation.

The primary movements associated with the upper extremities are shown in Figure 26-1. Most action involves two or more joints in simultaneous motion in three planes of space. Each type of joint is different and contains articular contours and traversing ligaments that determine the type of joint movement. Right- and left-side joints

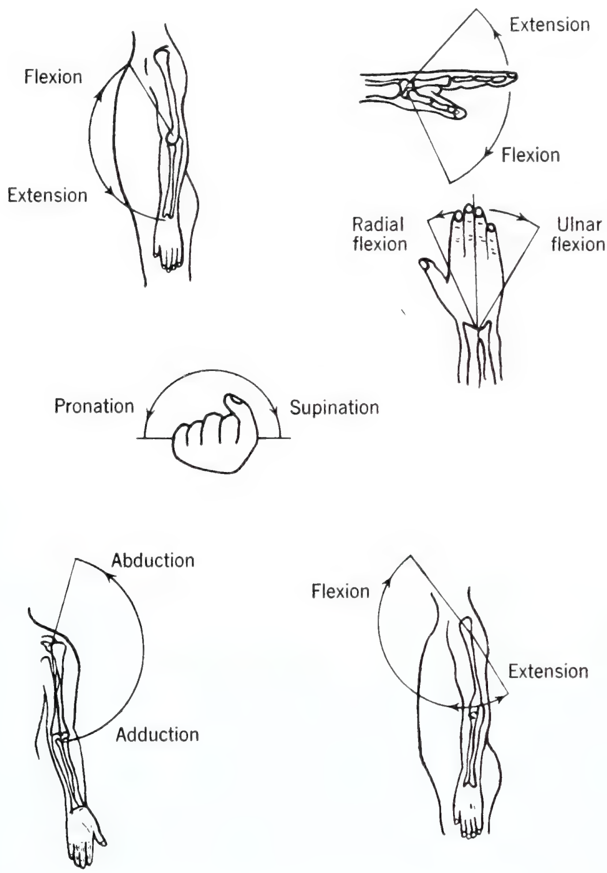


FIG. 26-1 *Movements of the upper extremities.*

represent essentially mirror images from the constructional viewpoint but usually differ to a small extent with regard to range of motion.

The quantitative kinematic description of the possible displacement of upper-extremity motion may be considered in a system of spherical coordinates. Dempster [4] has presented an excellent work in this respect, representing these movements in terms of globographic presentations wherein the angular range of joint motion is depicted on a globe with meridians and parallels. In this type of record a member of the joint is held fixed, with the functional center of the joint corresponding

to the center of the globe. The movable member is then used to describe a path that encompasses all its possible positions.

Kinetic description of upper-extremity motion is also complicated by the necessity for determining the physical parameters, such as mass, center-of-gravity location, and mass moment of inertia, of the various upper-extremity body parts. At present only approximate methods have been employed, involving comparisons of data accumulated from cadavers. Information of this nature is also presented by Dempster [4] and by Fischer [5].

A variety of photographic and electrical techniques are employed in measuring the kinematic and kinetic factors involved in the accomplishment of an upper-extremity task. Some of the basic optical techniques for recording movement are shown in Figure 26-2.

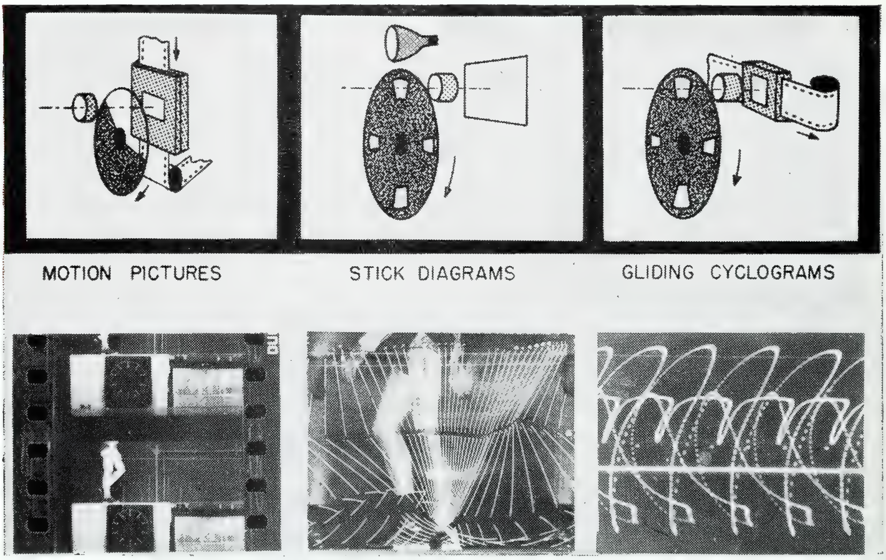


FIG. 26-2 *Optical methods of recording movements. (After R. Contini, R. Drillis, and L. Slote, Development of Techniques for the Evaluation of High Altitude Pressure Suits, USAF WADC Tech. Rept. 58–641, December, 1959.)*

FITTING AND HARNESSING

With the advent of externally powered prostheses, many features of the fitting and harnessing necessary for successful utilization of current prostheses will prove to be inadequate. Figure 26-3 shows a typical contemporary harness setup for a unilateral above-elbow amputee. The harness serves two basic purposes, providing suspension for the prosthesis and, more important, providing a source of power transmission for

operating the prosthesis's functional components. In this instance, residual body motions supply the power. Serious restriction of function is inherent in such a system because of the limitation of the motions that can be transmitted to this type of harness, lost motion within the socket, and force-transmission inefficiencies.

In general, the harnessing techniques that are at present being employed represent numerous compromises in connection with comfort,

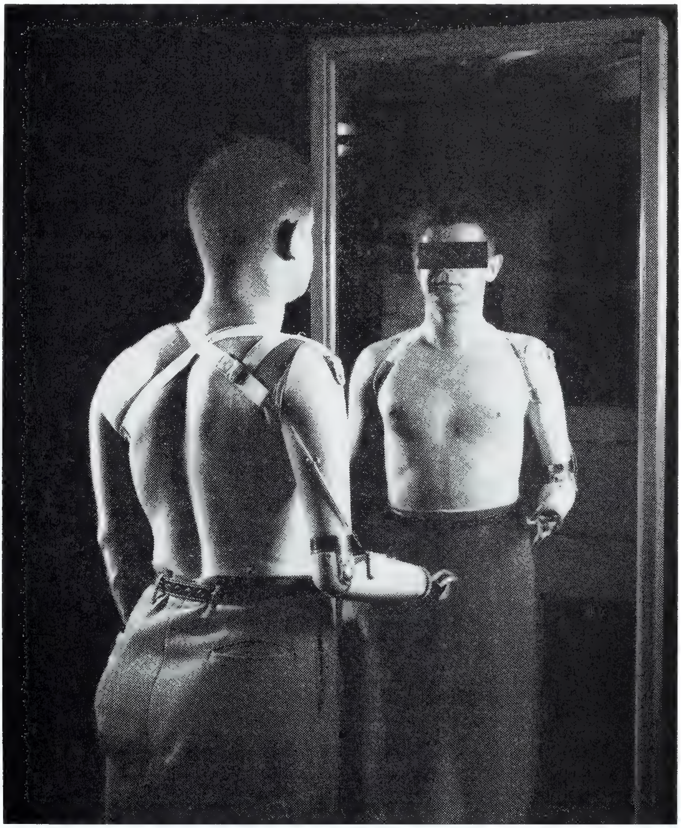


FIG. 26-3 *Typical above-elbow prosthesis.*

range of motion, and stability. More adequate stability between the prosthesis and the amputee is still needed in order to eliminate lost motion and provide more adequate control and feedback. The control pickups and associated sensory-feedback mechanisms of the externally powered prosthesis must have definite relations to body parts in order to ensure consistency of operation. Increased freedom of motion is required in order to permit operation of the prosthesis for combing the hair, reaching above the head, reaching behind the body, etc. Of equal

importance is comfort in the fitting process, since the device, no matter how well it functions, will not be useful if it causes intolerable pain to the wearer.

Noteworthy for use with current upper-extremity prostheses, as well as with externally powered versions, are some of the new techniques for achieving increased socket stability that have been developed in Germany [7]. These methods provide a more intricately shaped socket fitting, modifications of the socket for permitting increased range of motion, and detailed training techniques that aid in increasing prosthesis function.

Numerous surgical techniques are also being developed for providing a stable point of attachment for the prosthesis. For example, the cineplasty technique [2], which is already in use, utilizes a tunnel that is surgically formed in various muscles of the body (Figure 26-4). These

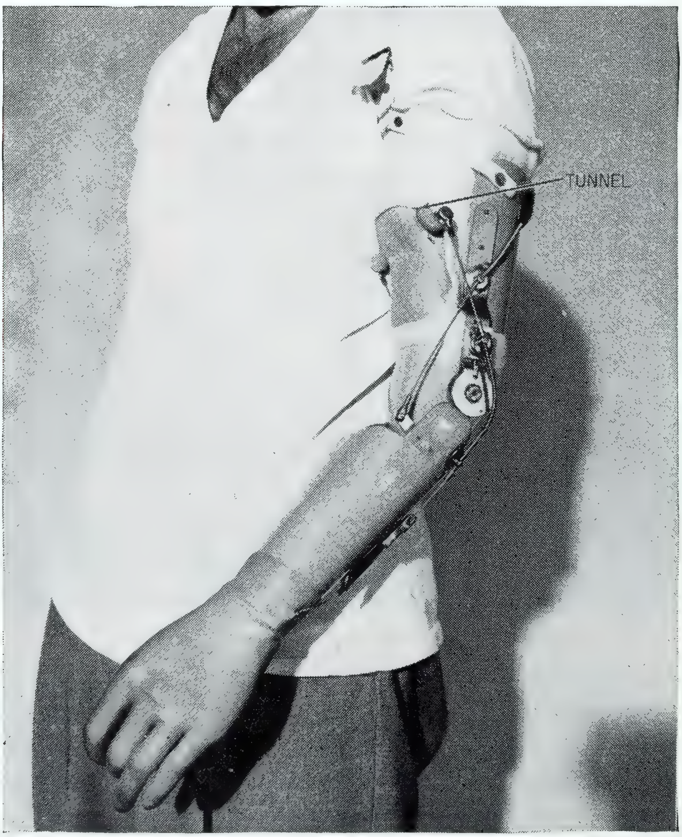


FIG. 26-4 *Biceps-muscle cineplastic tunnel. (From A. Staros, Some Human Factors Considerations in the Development of Upper Extremity Prostheses, ASME Rept. 59-A-184, 1959. By permission of the publishers.)*

tunnels provide a certain amount of stability for the prosthesis and a convenient connection to an excellent source of power, control, and feedback by tying into an integral part of the body.

The possibility of direct skeletal attachment has also been considered. Fundamental research in this field has been under way for some years, but no significant progress has been reported in relation to the extremities. Concurrent with research of this nature is the utilization of semiburied implants of inert material within the body. This has resulted in limited success with movable artificial eyes, implant dentures, and experimental devices in animals.

The use of chemical adhesives for securing the prosthesis to the body is another avenue of investigation. At present there are a number of adhesives that do not cause adverse reactions when placed in contact with the skin. For example, adhesives are currently employed in securing purely cosmetic restorations such as ears and noses, and the use of Eastman 910 cement has been successfully employed in eliminating sutures by cementing the skin together. Further development along these lines could prove fruitful and provide for a better fixation of the prosthesis to the skin.

CONTROL SOURCES

Inherent in the process of designing an upper-extremity externally powered prosthesis is the consideration of methods and techniques for enabling the amputee to exercise control over the prosthesis. The source of power will, of course, influence the specific circuitry of the control system and, to some extent, the type and location of the control site. As in the case of feedback sources, considerations of human capabilities and tolerances are required for the effective determination of source locations and the quantity of controls that can be implemented by the amputee.

One control source of promise consists in the bioelectric potentials that exist within human tissue. It is known that stimulation of nerve fiber results in impulses that cause the muscle to contract. This muscle action is manifested by the appearance of electrical activity at specific frequencies. Details of this process are covered in various medical publications, along with schemes for utilizing this phenomenon in conjunction with prosthetics [8, 11]. Problems still exist in effectively isolating the signals, identifying patterns of the signal for a given function, determining signal regeneration times, etc. At this point it suffices to say that the electromyographic (EMG) signal properties inherent in the human control system should in the near future be capable of providing excellent

sources of control for powered prostheses. Simultaneously the engineer will be faced with problems of devising methods and equipment to enable proper utilization of these currents to effect adequate wearer control over a powered device. A Soviet example of the utilization of this principle is given in Figure 26-5.

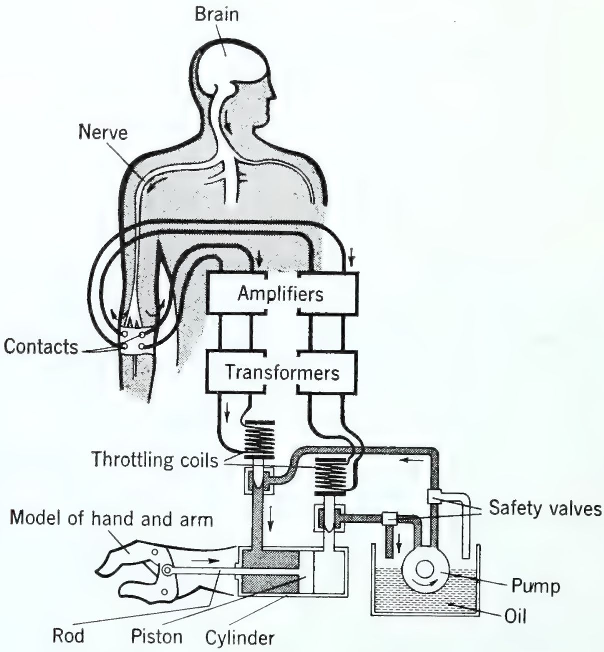


FIG. 26-5 *Russian electromyographic control arm.*

With the use of the cineplasty technique in conjunction with powered prostheses, control function can be obtained by means of low-force limited-excursion movements to activate the power source. Small, isolated muscle tunnels and tendon loops throughout the body may be utilized for the control of independent actions.

Numerous other surgical techniques are envisioned that would effectively employ portions of the body for control purposes. For example, the creation within the muscle itself of closed-end pockets in which a squeeze-type pressure transducer might be surgically implanted could result in another control source based on muscle action. In connection with this technique, consideration has been given to an embedded sensing device that utilizes a magnetic field to transmit a signal to an external pickup. Information could thus be obtained through the unbroken skin without the necessity of the protruding wires required for the normal

electrical transducer. Another technique isolates small sections of the muscle to provide an isolated control bulge. With proper training the amputee should be capable of exercising this type of control apart from other gross motions.

Whereas the previous discussion has focused on some of the more "exotic" methods for implementing the control function, there are routine procedures that are currently employed in self-powered prostheses and that could possibly be used in externally powered versions as well. For example, more effective utilization of nudge and shrug controls currently employed could be brought about by increasing the stability of attachment between the body and the prosthesis. Also, by proper shaping of the socket, the properties of muscle bulge might likewise be used for control.

POWER SOURCES

In any study of externally powered upper-extremity prostheses and their design requirements, one must initially consider sources of power that can be utilized. There are specific requirements that make the choice among sources a rather complicated one. A few of the requisites are: (1) high efficiency with respect to ratios of energy to weight and volume; (2) facility for simple power regeneration (e.g., regeneration by household current); (3) high reliability; (4) low noise; (5) extended mechanical-life characteristics; (6) compatibility with a variety of systems, current and future; (7) sufficient storage capability for at least 1-day operation; (8) no adverse effects to the wearer; (9) immunity from adverse environmental conditions; (10) low operating cost.

Current practice employs body-power sources that are close to the prosthesis. Thus, by certain body motions of an upper extremity, it is possible to obtain power for some limited operations. Attempts have been made to utilize more complicated harnessing procedures so that various body gyrations could produce a number of additional motions. For the most part, however, these more involved endeavors have proved futile because the resulting cosmesis, amount of concentration required of the amputee, and reduced efficiency have neutralized any gains in mobility.

The use of remote body functions for providing power has also been given some consideration. Hydraulic methods have been adapted to provide for the transfer of this body power to the prosthesis. An example of such a system is a hydraulic cylinder operated by the foot, with fluid lines running up the body to the upper-extremity prosthesis. Opposition to this form of technique is based upon the necessity for extensive hydraulic circuitry, with its adverse effects to the wearer in terms of

comfort and cosmesis. Nevertheless, although the use of hydraulics for transferring power does not constitute an external-power function, it does point up the flexibility of external-power utilization. For example, when the upper-extremity prosthesis is not in use, external body parts may be used to regenerate an external-power source (e.g., recharge a battery or build up pressure in a cylinder) during normal activities such as walking.

To date, consideration of true external-power sources has been limited to the use of electricity and compressed gases. Electricity offers the advantages of efficient energy storage and practical weight and volume. In addition, it offers compatibility with the control and feedback capabilities of the human nervous system. Recent developments in the miniaturization of circuitry can well be applied in the design of prosthetic devices. Bernshtein [3] has also indicated that only three electric motors are necessary to accommodate the numerous degrees of freedom required in prosthetic motion.

The electric arm [1] is a recent example of the application of electric power to prosthetics. This device employs a bidirectional electric motor that transmits power to the various components through electrically operated clutches. Hand prehension, wrist rotation, wrist flexion, upper-arm rotation, and shoulder flexion can be provided by the various models. Power for the system was derived initially from lead-acid cells and subsequently from silver-zinc cells. All cells were rechargeable. Nevertheless, numerous mechanical breakdowns were characteristic of the arm's operation, and at present use of the device is discontinued. Figure 26-6 illustrates a below-elbow model of this arm.

Actuators utilizing compressed gases, such as carbon dioxide, Freon, hydrogen peroxide, and liquid oxygen, appear promising because of their light weight. Several prostheses have been developed utilizing carbon dioxide, and this gas seems to be preferred because of both its availability and its low cost. In a typical pneumatic prosthesis using carbon dioxide, the compressed gas is stored in a portable tank containing sufficient carbon dioxide to provide a unilateral amputee with approximately 8 hr of prosthetic application. The gas is stored at a pressure of 850 psi at 70°F and is then fed to a control regulator that reduces it to 80 psi, the operating pressure. Double pneumatic cylinders are employed and are activated by a series of control valves.

Integration of several sources of power such as gas and electricity is also feasible. Currently, however, it appears that electrical power will play the major role in future externally powered upper-extremity prostheses, for it is most compatible with the human nervous system and can be utilized most readily in conjunction with human control and feedback functions.

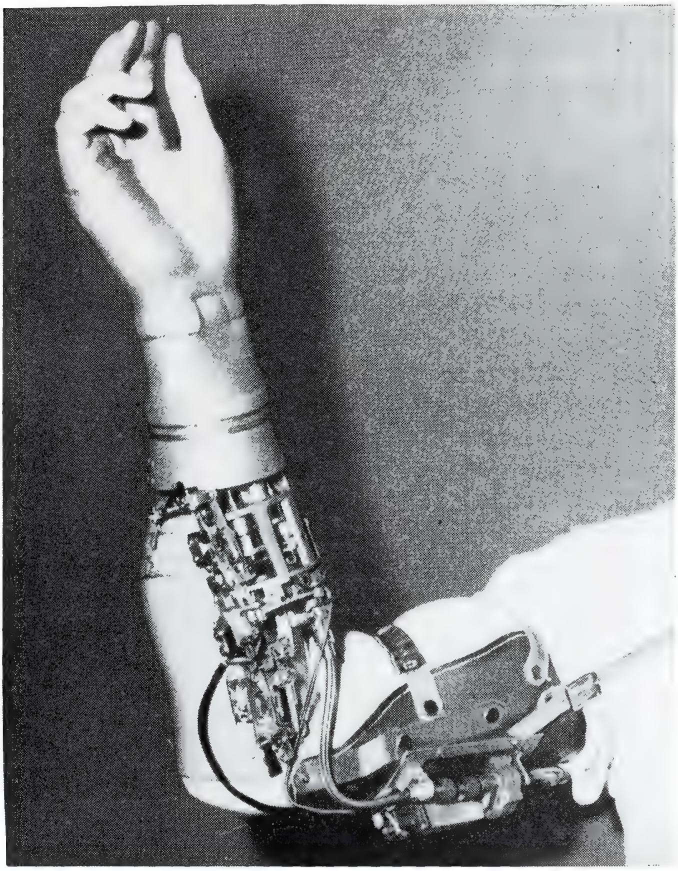


FIG. 26-6 *IBM-Alderson below-elbow electric arm.*

FEEDBACK SOURCES

Consideration must be given to providing the amputee with knowledge of the position of the prosthesis and the force outputs at the terminal device. In present practice, vision is relied upon for prime information on position, and it is logical to assume that vision will also play a major role in powered prostheses. With regard to force feedback, the present harness system provides some indication of body-force output. However, there are many inefficiencies inherent in this type of system. Moreover, in a powered prosthesis the harness will probably be divorced from the power source and thus will not be capable of transmitting such information.

The cineplasty technique provides for an excellent indication of

force output, for it involves direct attachment to muscles capable of full signal transmission to the brain. Therefore, a feedback source of this type can be used for providing force cues in a powered prosthesis. In addition to supplying force feedback, the muscle tunnels can accomplish a control function by providing small excursions and forces for activating a switch or valve.

The use of skin senses [6] can also provide feedback. Heat, cold, and pain have properties that make these sensations unsuitable as communication sources, but the application of touch or pressure to various body sites appears to have some usefulness. One disadvantage is that the body readily accommodates to sustained, tolerable pressures, and therefore the individual's response is limited.

A vibratory pressure stimulus is more effective as a means of communication. The variables that can be used in the application of vibratory pressure are the area, intensity, frequency, and duration of the stimulus. With information about these dimensions, suitable codes can be formulated that permit transmission of the required information. By proper training and conditioning the amputee should have little difficulty in being able to discern the various signals and interpret their significance. Using this technique will probably require development of transducers that will not have appreciable damping when in contact with the skin. It may be possible, however, to bypass the transducer and provide electrical vibratory stimulation directly to the skin.

OTHER FACTORS

In addition to the major problems of fitting and harnessing, control sources, power sources, and feedback sources, it is necessary to consider the psychological and safety factors, cosmesis, comfort, and maintenance, that are involved in the relation between the amputee and the prosthesis. Although seemingly secondary to operational aspects, these factors are important because they, too, determine whether or not the prosthesis will be worn.

Cosmetic aspects include the size, weight, shape, and noise factors of the prosthesis. It must be remembered that the device in its final form should have the same shape as the normal upper extremity and should conform closely to the sizes of the nonamputated side or, in the case of a bilateral amputee, to the general dimensions of the body. Weight is another important factor, for the amount of bodily energy required for transporting an overheavy device would seriously limit the effective utilization of the prosthesis.

It is also necessary to minimize noise effects during operation of the

prosthesis. The whirring of gears, the noises associated with the opening and closing of valves, and the click of an engaging clutch all tend to diminish the amputee's acceptance of a prosthesis.

Comfort may be improved considerably by distributing forces over as wide an area as possible. In addition, the choice of proper construction materials can influence the amputee-prosthesis relation by providing for cooling (e.g., by means of porous laminate), pressure distribution, and the absence of skin reactions. It is mandatory that any device also employ the fail-safe concept; the device must be constructed in such a way that no harm is done to the amputee by the prosthesis. For example, in the cineplasty technique, attachment to the muscle tunnel employs a cable with a breaking strength well under that of the tunnel; thus damage to the skin or muscle tissue is prevented. It is advisable also that duplicate circuitry be provided in the prosthesis in order to ensure operation in emergency situations.

Maintenance difficulties can cause amputee rejection as well as minimize effective prosthetic restoration; the prosthesis that is continually in the "shop" is not useful. A prosthetic device should require ideally no more than annual routine inspection and lubrication.

CONCLUSION

The externally powered prosthesis is a complex machine involved in a tight psychological, physiological, and technological relation to the user. Some of the problems in this man-machine system demand atypical design procedures. According to Weiss [10], the more complex the device, the better the chance for amputee acceptance, since a complex prosthesis is a closer approximation of the complicated anatomical structure that has been lost.

Consideration must be given to the control-force-feedback relation that must be maintained in order to utilize effectively the remaining bodily systems. Further quantitative investigations of decision making, design of controls, processing of information, man-machine dynamics, and servo analysis are still required in the development of amputee-prosthesis systems. The use of surgical techniques to enhance operation of the system should be encouraged. In all probability future prosthetic systems will utilize electronic circuitry and electrical power tied into the bioelectric control and feedback systems contained within the body. Therefore, special attention should be given to the developmental problems in this area.

The solution of the problem of functional upper-extremity restoration obviously is not now at hand but lies in a future crowded with advance-

ments in medicine and technology. It is also possible that future biological advances in the regeneration [9] and transplant of limbs may even abrogate the need for mechanical prostheses.

BIBLIOGRAPHY

1. Alderson, S. W.: The Electric Arm, in P. E. Klopsteg et al. (ed.), "Human Limbs and Their Substitutes," pp. 359–408, McGraw-Hill Book Company, Inc., New York, 1954.

2. Bechtol, C. O., and G. T. Aitken: Cineplasty, in "Artificial Limbs," vol. 2 of "Orthopaedic Appliances Atlas," pp. 437–449, J. W. Edwards, Publisher, Inc., Ann Arbor, Mich., 1960.

3. Bernshtein, N.: Analysis of Some Spatial Movements of the Arm and Proposal on Construction of a Work Prosthesis, *Trans. Central Inst. Prostheses,* vol. 1, 1957. (In Russian.)

4. Dempster, W. T.: Space Requirements of the Seated Operator, *WADC Tech. Rept.* 55–159, 1955.

5. Fischer, O.: "Theoretische Grundlagen für eine Mechanik der Lebenden Körper," Teubner, Verlagsgesellschaft, Leipzig, 1906.

6. Hahn, J. F.: Communication via the Skin, in The Application of External Power in Prosthetics and Orthotics, *Natl. Research Council Publ.* 874, pp. 145–149, 1961.

7. Hepp, O., and G. G. Kuhn: Upper Extremity Prostheses, in Prosthetics, pp. 133–181, International Committee on Prostheses, Braces and Technical Aids, International Society for the Welfare of Cripples, Copenhagen, 1960.

8. Kobrinskiiy, A. Y.: Bioelectrical Control of Prosthetic Devices, *Herald Acad. Sci.* (U.S.S.R.), vol. 30, no. 7, July, 1960. (*Office Tech. Serv. Transl.* 61 11074 B 50.)

9. Singer, M.: The Regeneration of Body Parts, *Sci. American,* vol. 199, no. 4, pp. 79–88, October, 1958.

10. Weiss, S. A.: The Body Image as Related to Phantom Sensation: A Hypothetical Conceptualization of Seemingly Isolated Findings, *Ann. N.Y. Acad. Sci.,* vol. 74, pp. 25–29, September, 1958.

11. Weltman, G., et al.: An Analysis of Bioelectric Prosthesis Control, *Univ. Calif. Los Angeles Biotechnol. Lab. Tech. Rept.* 1, 1959.

Chapter 27

MANIPULATORS USED FOR HANDLING RADIOACTIVE MATERIALS

*R. C. Goertz**

OPERATIONS INVOLVING RADIOACTIVE MATERIALS ARE USUALLY CARRIED OUT within shielded enclosures. These enclosures usually have walls that are several feet thick and made of dense concrete; sometimes steel or lead is used. The operator must stay outside these enclosures when intense, penetrating radioactivity is in the enclosure and not otherwise shielded.

Remotely controlled devices must be provided for carrying out the experiments or operations within the shielded enclosures. One of the most interesting and complex of these remotely controlled devices is the general-purpose manipulator. This manipulator is designed to perform nonroutine handling and manipulations that would ordinarily be performed directly with the hands. The general-purpose manipulator consists of a mechanical arm capable of grasping an object and moving it to any translational and angular position within the working volume of the manipulator. The object may be of almost any size, shape, and weight within the capacity of the manipulator. The manipulator arm usually has a minimum of seven independent motions. These motions are remotely controlled by a human operator who is usually just outside the shielding wall and in front of one of the large shielding windows.

The general-purpose manipulator may be used for moving objects, moving levers or knobs on apparatus, assembling parts, and manipulating wrenches. In all these operations the manipulator must come into physical contact with the object before the desired force and movement can be made on it. A collision occurs when the manipulator makes this contact. General-purpose manipulation consists essentially in a series of collisions with unwanted forces, the application of wanted forces, and the application of desired motions. The collision forces should be low, and any other unwanted forces should also be small.

* Argonne National Laboratory, Argonne, Ill.

This chapter is based on work performed under the auspices of the U.S. Atomic Energy Commission.

General-purpose manipulators have been developed in two distinctly different designs. One of these designs utilizes a unilateral electrical drive and control system; that is, the handles that are controlled by the human operator do not receive feedback from the mechanical arm. The other type of design is called a master-slave manipulator and utilizes a bilateral mechanical or bilateral electrical control system. The bilateral control system feeds back force and position information to the handle that is controlled by the human operator.

Before describing and comparing these two designs, a brief review of some of the basic characteristics of general-purpose manipulators is necessary. The general-purpose manipulators to be described have all their motions controlled by a human operator. The only exceptions are those controls which limit positions and, sometimes, maximum forces. These manipulators also have a mechanical input and output. The mechanical input is the force and movement of the handle or handles that the human operator manipulates with his hands. The principal mechanical output is the force and motion that are imparted by the manipulator arm to the object. In the master-slave type of manipulators there is also feedback to the control handle of the forces and motions acting on the slave arm. In addition there have been a few manipulators that have utilized audible or visual aids to indicate the gripping force of the tongs.

All general-purpose manipulators must have at least seven independent motions. These motions must include three for translating the object from place to place, three for rotating the object in space, and one for grasping the object. Figure 27-1 illustrates these motions by showing an object in the coordinates X, Y, and Z with the angular coordinates a, b, and c. The simplest manipulators have only these seven independent motions. Additional motions are sometimes incorporated to change the shape of the arm or to provide additional finger movements. All the manipulators described in this chapter have these seven independent motions, with the exception of one that has an additional movement of the lower arm at the shoulder joint.

The basic mechanical arrangement illustrated in Figure 27-2 is often used for manipulators that are controlled with unilateral electrical devices. The seven independent motions are achieved by utilizing cartesian coordinates for the translational motions. The X and Y motions are provided by the bridge and rail system, the Z motion by the telescoping boom, and the rotational motions by the wrist joint at the lower end of the boom. The tongs, or gripping device, are attached to the end of the wrist joint. Usually each of these seven motions is equipped with an electric motor that drives the particular motion at a predetermined speed after the operator has closed the switch for that particular motor. The

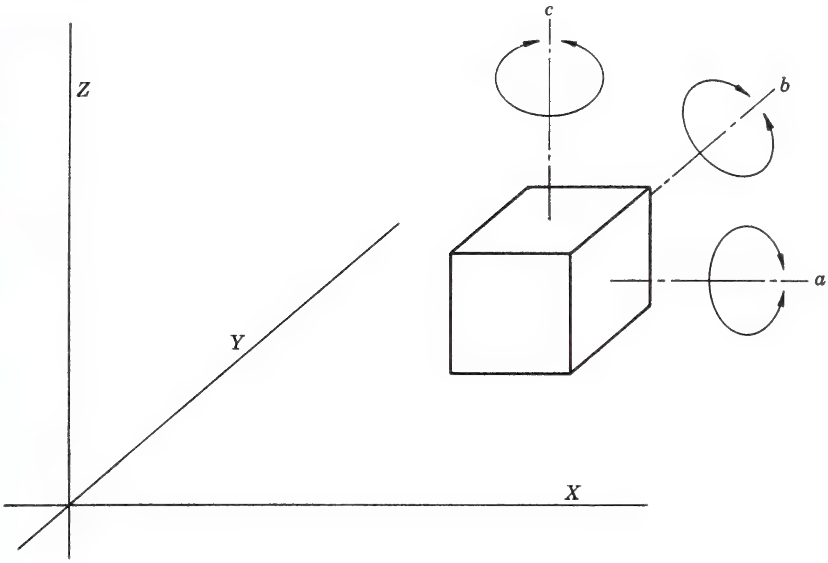


FIG. 27-1 *Solid object in space.*

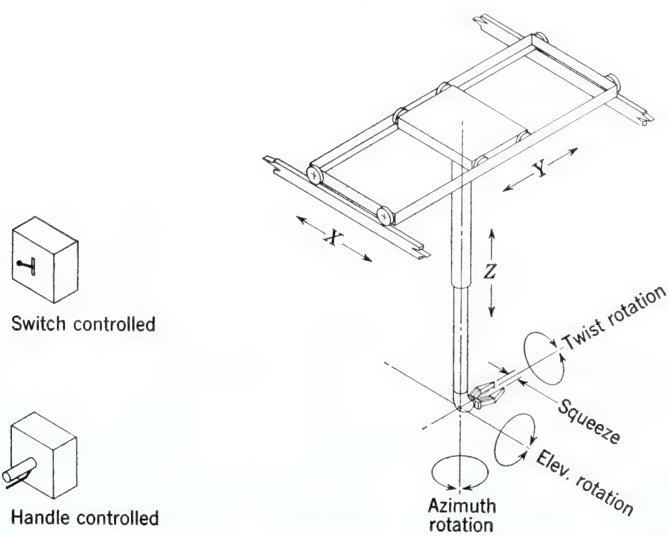


FIG. 27-2 *Rectilinear manipulators with unilateral electrical control.*

switches are usually of the on-off-on spring-centered type. Movement of the switch in one direction supplies a voltage having a particular polarity; movement in the opposite direction supplies the same voltage with the opposite polarity. The switches may be operated individually, or several of them may be mechanically connected to one control handle.

A variation in this type of control involves the substitution of proportional-signal devices for the switches. These devices provide an electrical signal that is proportional to their displacement from the zero position, with a polarity corresponding to the direction in which the device is moved. This signal then goes through a suitable amplifier system, and the drive motor of the particular motion is made to turn at a speed proportional to the voltage. Thus the speed of a particular motion of the manipulator is made proportional to the displacement of the control handle. The use of a proportional-signal device by comparison with the use of a switch gives much smoother control of the manipulator motions but still does not control forces.

Figure 27-3 is a line drawing of an actual manipulator utilizing the unilateral switch-control system. Each of the seven d-c motors is controlled by an individual switch. These switches are grouped together in a small control box, which is shown in front of the window and also in an

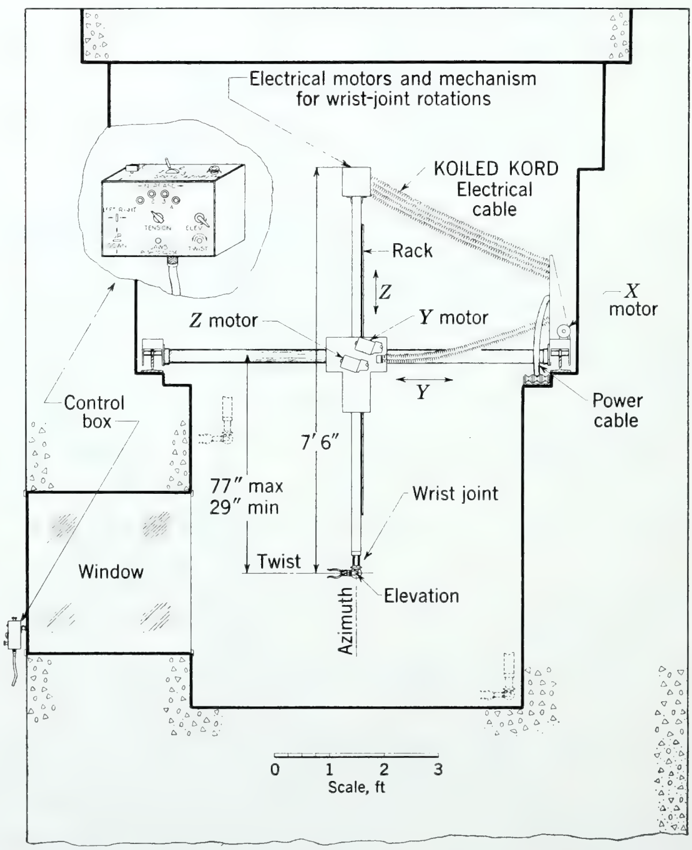


FIG. 27-3 *Rectilinear electric manipulator, model 4, with unilateral electrical control.*

enlarged isometric view. All the seven switches are of the on-off-on spring-centered type. The X, Y, and Z control switches are located on the left portion of the control box and are oriented so that the manipulator moves in the direction in which the switch is moved. In the isometric view, they are labeled LEFT-RIGHT, IN-OUT, and UP-DOWN for X, Y, and Z, respectively. The azimuth switch, a spring-centered knob, is located on the top and at the right of the control box. The azimuth motion turns in the direction in which the knob is turned. The four switches, X, Y, Z, and azimuth, are oriented so that movement of the switch will cause the manipulator arm to move in the same direction.

The other three motions do not have a simple correspondence of manipulator movement with switch movement. The elevation switch has a handle pointing down about 45°, partially simulating the position of the tong axis that moves from the horizontal position, as shown, to the vertically down position. The twist-rotation control switch is a knob on the lower right of the control box. If the tongs are oriented as shown, they will rotate (twist motion) in the direction that the knob is turned in. If, on the other hand, the tongs are pointing in the opposite direction, then the tongs will rotate in a direction opposite to that in which the knob is turned. Closing or opening the tongs is accomplished by actuating a push-pull switch with a spring-centered off position. This switch also usually has no motion correspondence.

A person with good manual dexterity and with mechanical experience can learn to operate the control devices no matter how they are located or oriented. However, the learning period is considerably longer and the frequency of making mistakes is considerably higher in the operation of controls that do not have a simple correspondence of their movements with the resulting manipulator movements.

There is no force feedback to the controls on this manipulator. The operator must depend upon his visual perception of the deflections of the manipulator arm or of the apparatus after the manipulator has come into physical contact with the apparatus. The gripping force, however, can be preselected within limits.

Another rectilinear electric manipulator incorporating unilateral controls is illustrated in Figure 27-4. This manipulator has a control box that is equipped with two handles. One of the handles operates all the switches for the translational motions, and the other handle operates all the switches for the rotational motions. The only feedback from the manipulator arm to the control box is a signal that causes an electric meter to move in proportion to the squeeze force of the tongs. This manipulator has eight motions. The additional motion is the movement of the lower arm at the shoulder joint, which increases the ability of the manipulator arm to reach in horizontal directions.

Many different sizes of manipulators have been designed with mechanical arrangements and unilateral control systems similar to the arrangements in Figures 27-3 and 27-4. They vary in maximum load capacity from about 5 to 500 lb. The larger manipulators have the advantage of being able to lift and exert larger forces than can human hands and to act as a crane as well as a manipulator. However, the opera-

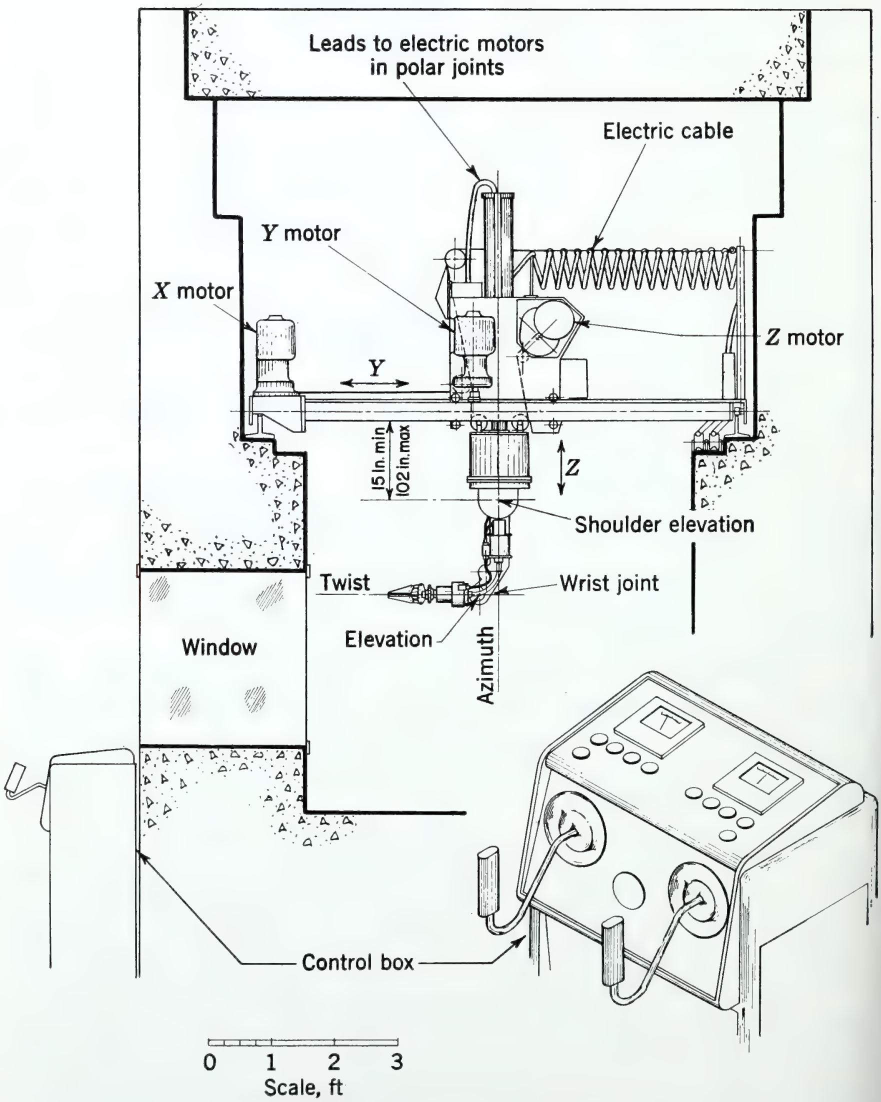


FIG. 27-4 *A commercial rectilinear manipulator with unilateral electrical control.*

tion of these manipulators is slow and awkward compared with that of master-slave manipulators.

The other type of general-purpose manipulator, the bilateral, or master-slave, manipulator, is illustrated in its basic arrangement in Figure 27-5. The master-slave manipulator has two similar arms. One of the arms,

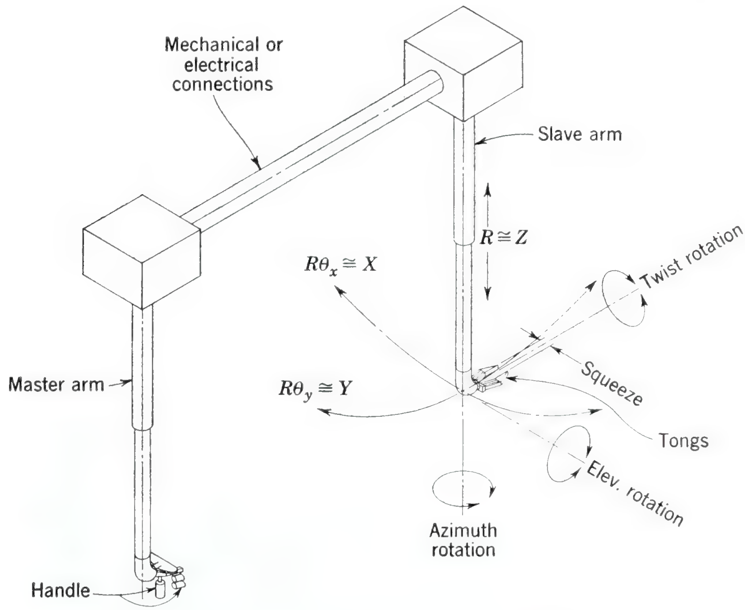


FIG. 27-5 *Basic arrangement for some master-slave manipulators.*

the slave arm, is located inside the shielded enclosure; it is equipped with tongs for grasping objects. The other arm, the master arm, is located outside the shielded enclosure and is equipped with a handle by means of which the operator can produce any or all of the arm's seven motions at one time. Each of the seven motions of the master arm is connected to the corresponding motion of the slave arm. The connecting linkages are mechanical or special force-reflecting electric servos. When the master arm is moved, the slave arm moves a corresponding amount. The arms are well counterbalanced, and all motions operate with low friction. The operator moves the master arm in the direction in which he wants the slave arm to move.

Figure 27-6 is a line drawing of the master-slave manipulator called model 8. This manipulator is at present used in most shielded facilities in this country and in many facilities in other countries. It is shown mounted in a shielding wall about 3 ft thick. All seven of its independent motions are connected from the master arm to the slave arm by stainless-

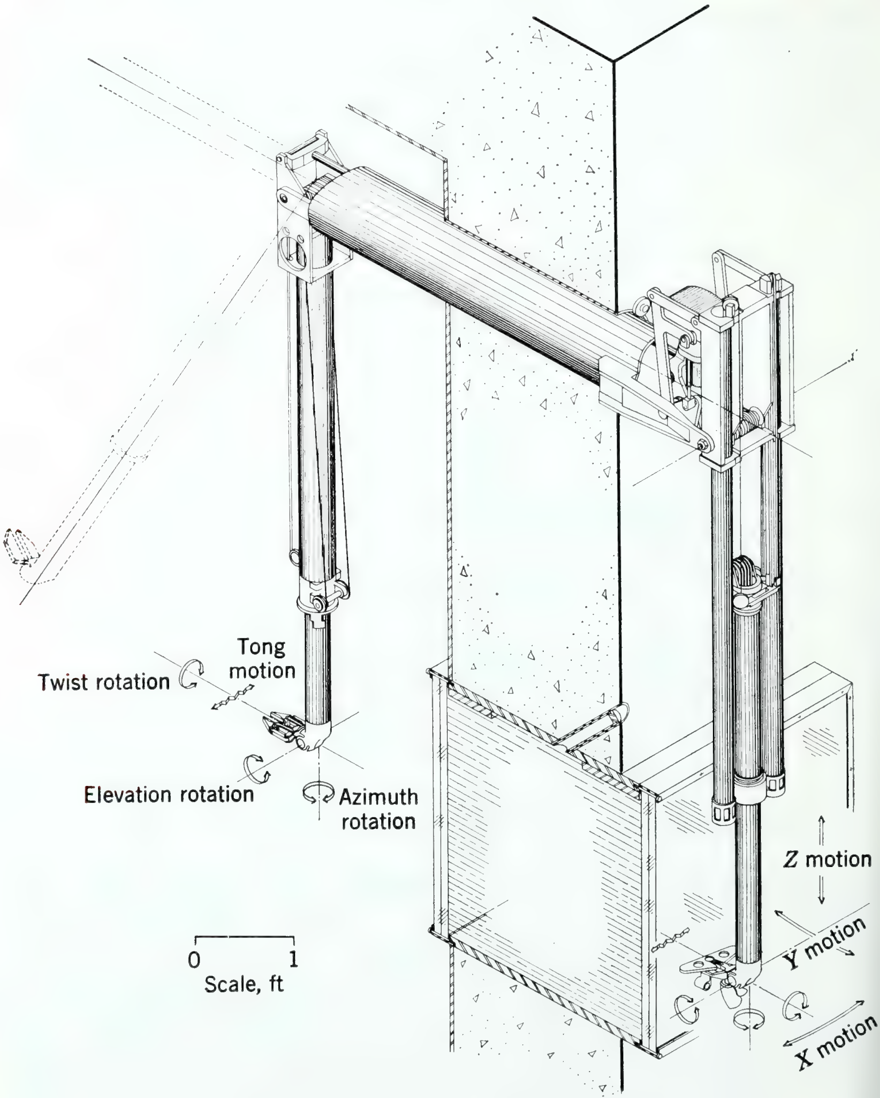


FIG. 27-6 *Master-slave manipulator, model 8.*

steel cables, Elgiloy[1] tapes, and other mechanical linkages. These linkages cause the slave arm to move the same amount as the master arm and in the same direction. The friction in all the motions is kept reasonably low by using ball bearings for all moving parts except for a few parts in the

[1] Elgiloy is the trade name of a special corrosion-resistant alloy that has a yield strength of about 250,000 psi when in its half-hard condition.

handle and in the tongs. Figure 27-7*a* and *b* gives detailed pictures of the handle and tongs used in the model 8 manipulator. These have also been adapted for use with other master-slave manipulators.

The connecting linkages not only cause the slave arm to follow the master arm but also transmit back to the master arm the load forces acting against the slave arm. The operator feels these loads reasonably

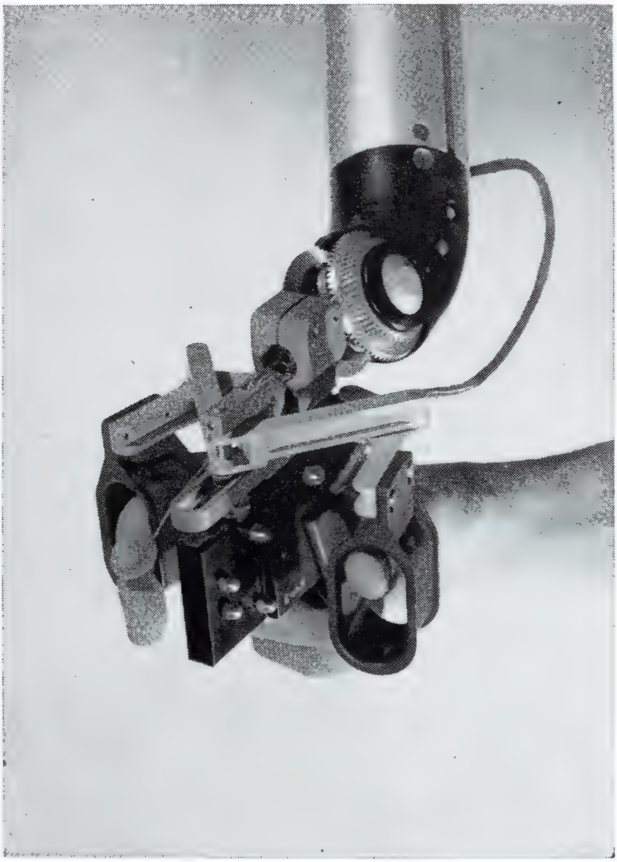


FIG. 27-7*a* *Handle used in the model 8 manipulator.*

well but with some masking effects due to friction, mass, and springiness. All the motions are mechanically reversible or bilateral; that is, if the slave arm is moved, the master arm will follow it. Thus the operator can cause the slave arm to perform operations that would be either extremely difficult or tedious to do or could not be done with manipulators utilizing unilateral drive and control systems. For example, it is possible to turn a crank with the master-slave manipulator. The operator simply exerts a

force in the approximate direction that he wants the crank to move in; then the other five motions (excluding the tong squeeze, because it is locked onto the crank handle) are driven by the slave arm following the movement of the crank itself. Such responsive performance of the slave arm to the restraints imposed by the load is usually needed whenever an apparatus is being operated. The knobs, levers, and other parts of an

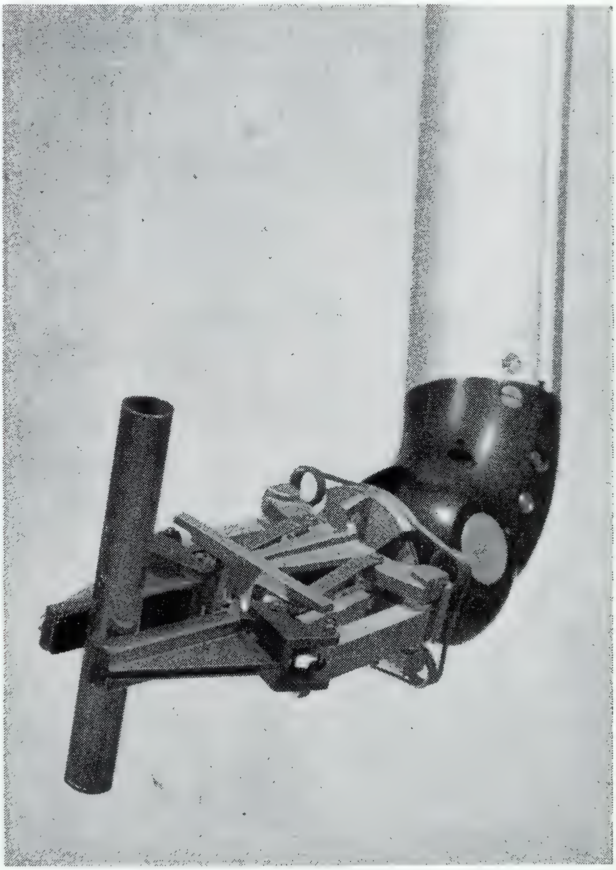


FIG. 27-7*b* *Tongs used in the model 8 manipulator.*

apparatus do not have the full six degrees of freedom of a free body. Usually they have only one degree of freedom and may be delicate enough so that they would be damaged if a large force were applied to them in one of the other directions.

These manipulators are usually provided and used in pairs and are usually mounted about 30 in. apart. One person operates both manipulators simultaneously, one manipulator with the right hand and the

other with the left hand. Since the manipulators are bilateral in all their movements, one manipulator can help the other; e.g., their combined forces can be applied to an object, or one can help the other to get a new grip on the object. They do not tend to fight each other as would be likely to be the case with unilateral drives and controls.

The load capacity of the model 8 master-slave manipulator is about 20 lb in any direction with the object held in the tongs, in contrast to a capacity of 5 to 500 lb for the unilateral drive and control manipulator. The average operator, however, cannot easily lift or exert a force much above 10 or 15 lb when he is operating the manipulator with only one hand. This limitation is due primarily to the poor coupling of the handle to the human hand. The operator often uses both hands for handling objects heavier than 10 lb. A handle is now being offered with an extension on it so that the operator can conveniently use both hands. The straight lifting capacity of this manipulator is about 100 lb if a suitable hook is attached just above the slave-arm wrist joint.

It should be noted that the human operator supplies all the energy for performing the manipulations. He also must supply the energy and forces necessary to overcome the friction and inertia of the manipulator. For operations dealing with weights under about 2 lb, operator fatigue comes primarily from the friction, mass, and awkwardness imposed by the manipulator and only to a small extent from the loads reflected from the object. When the loads are heavy, however, the reverse is true.

This master-slave manipulator is easy to operate. Nevertheless, the operator is required to have reasonably high manual dexterity and considerable experience in working on mechanical devices. Such a person can learn to operate the manipulator to a fair level of performance in less than 1 hr. High proficiency takes several months of experience. The operator soon learns to think of where he wants the tongs to move and not to think of the hand movements required to achieve the desired results. Later, he thinks principally of what operations he wants to perform on the object. This ease of learning is made possible by the manipulator's moving in correspondence with the movements of the operator's hands. Inherent and learned manual dexterity seems to work naturally even though the tongs that are performing the work may be 5 or 10 ft from the operator's hands.

Of course, the dexterity of the man-manipulator system falls far short of that in direct manual manipulation. One of the handicaps of the manipulator is that it has only seven independent motions. These are all the basic motions needed to move an object to any position or angular orientation; however, the seven motions are far fewer than are possible with the human arm and hand. The grasping operations have to be achieved by using the parallel, moving jaws of the tongs. In order to be reasonably

sure that the desired final movements are within the angular and displacement ranges of the manipulator, the operator must exercise considerable forethought before grasping an object. If the object must be later reoriented in the tongs, it has to be held by some other device while it is being reoriented or it has to be reoriented by another manipulator. Another limitation is brought about when the twist axis coincides with the azimuth, thus reducing the degrees of freedom to six. Again it is necessary to plan the movements so that this condition is avoided.

The master-slave manipulator, model 8, is counterbalanced in the X, Y, and Z motions to within a few ounces. This high degree of balancing is valuable because most of these manipulators do not have brakes to keep the arms from moving when the master arm is not being held. If brakes that were easily controlled at the master handle were provided, then there might be small advantage to the high degree of counterbalancing. It might even be advantageous to overbalance the Z motion in order to help the operator carry the loads. The wrist angular motions are not counterbalanced, since the weight of the handle and tongs is only a little over 2 lb, and there has been little complaint about the lack of counterbalance.

When the manipulator is in good working condition, the friction is about 2 oz in the X and Y motions and about 12 oz in the Z motion. The 12-oz friction in the Z motion is probably a little disturbing, but the operators have not complained. Some of the operators, however, have had experience on earlier, lighter-duty manipulators that had less friction, and they prefer the lower amount. If this friction increases to 1.5 or 2 lb, the operating personnel usually ask for the manipulator to be removed and repaired. The friction in the wrist-joint rotations is 18 oz-in. for elevation, 8 oz-in. for twist, and 21 oz-in. for azimuth. This friction does not cause noticeable trouble.

The mass of the manipulator, as felt by the operator, is somewhat as follows: For the X and Y motion the equivalent mass is 6.6 lb when the arms are fully extended and 18.8 lb when the arms are telescoped to the upper limit; the mass for the Z motion is about 16 lb. The inertia, or mass, of the wrist-joint angular movements is as follows: The elevation-motion inertia is 45 lb-in.^{2}, or 1.25 lb at the center of the tongs, the twist-motion inertia is 9 lb-in.^{2}, and the azimuth-motion inertia is 76 lb-in.^{2}, or 1.44 lb at the center of the tongs. These inertias do not seem to give any trouble.

The springiness of the manipulator can be measured by holding the slave tongs fixed, applying a force or torque at the master handle in the direction of one of the motions, and measuring the deflection at or near the master handle. This may be repeated for each motion. With a 20-lb force applied, the deflection is 2.5 in. in the X direction, 3.5 in. in the Y direction, and 0.125 in. in the Z direction. With a 20 lb-in. torque applied,

the angular deflection in the elevation motion is 3.5°, the deflection in the twist motion is 8°, and the deflection in the azimuth motion is 2°. Springiness, or sponginess, reduces the higher-frequency components of forces reflected back to the operator. This causes some difficulty in such operations as putting a hex wrench into a small (e.g., No. 8) hollow hex-head screw; it is difficult to "feel" when the wrench passes over the hole.

In spite of these limitations, a skilled operator can perform manipulations such as weighing, testing hardness, moving levers, making measurements, and so forth, at speeds of one-tenth to one-fourth the speed of direct manual manipulation.

Mechanically connected master-slave manipulators have a limited working volume, since they are anchored to the shielding wall. Because of this limitation, master-slave manipulators that have only electrical wires connecting the master and slave arms have been developed. The mechanical linkages then are replaced by special force-reflecting electric servos developed for this purpose. These servos cause the slave arm to follow the movements of the master arm and to reflect the load forces from the slave arm back to the master arm.

Figure 27-8 is a block diagram of one of these force-reflecting servos. There is a master-servo assembly and a slave-servo assembly, each consisting primarily of a gear box, a servomotor, a position-data device,

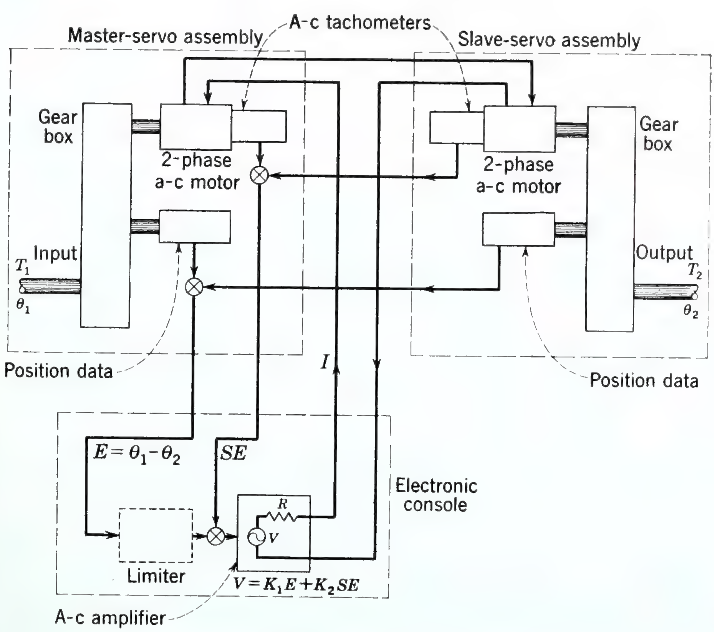


FIG. 27-8 *Diagram of simple force-reflecting servo.*

a tachometer, and an input or output shaft. Auxiliary to these assemblies is an amplifier system. The behavior of the slave servos is similar to that of an ordinary electric servo utilizing a two-phase motor and tachometer for viscous damping. The master-servo assembly provides the reaction force that is proportional to the load force on the output shaft. The error voltage E, which is proportional to the difference between the positions of the output and the input shafts, is passed through a limiter and into an electronic power amplifier. The tachometer voltages are mixed and also fed into the power amplifier. The output from this amplifier supplies power to the slave-servo and master-servo motors. The motors are connected in series and are of a polarity that causes the slave motor to drive the output in accordance with the input and to cause the master motor to give a reaction torque tending to oppose the mechanical input. The limiter, through which the error voltage passes, limits the error voltage to the amount required to give full output of the amplifier when there is no signal from the tachometer generators. If the servo output is considerably out of synchronism with the input, the tachometer generators, because of the limited error voltage, tend to limit the speed of synchronization. This slow synchronization is incorporated in order to protect the equipment at the slave arm or the human operator when the power is turned on or when the arms are forced out of synchronism by an overload.

Figure 27-9 is a simple diagram showing the mechanical equivalent network of the force-reflecting servo of Figure 27-8. The following

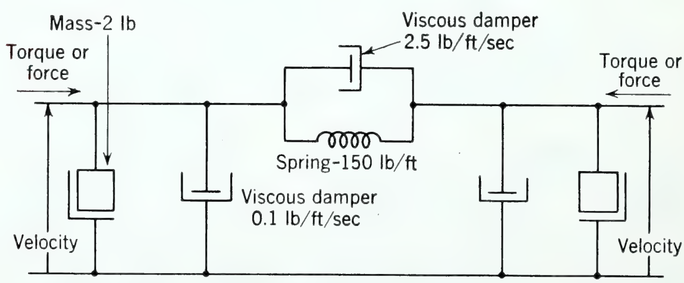


FIG. 27-9 *Mechanical schema of simple force-reflecting servo.*

constants are for the servo when it is incorporated into the master-slave manipulator, model 2, illustrated in Figure 27-10. The input mass of 2 lb represents the master-servo motor and gear train. The first-shunt viscous damper of 0.1 lb/ft/sec is caused by mechanical and electrical drag in the master-servo assembly. The series spring of 150 lb/ft represents the servo stiffness as measured for the X, Y, and Z directions at the wrist joints. The series viscous damper of 2.5 lb/ft/sec is the damping

caused by the tachometer generators and, to some extent, by the damping effect of one motor on the other.

The manipulator shown in Figure 27-10 incorporates one of the above servos in each of its seven independent motions. For the first test

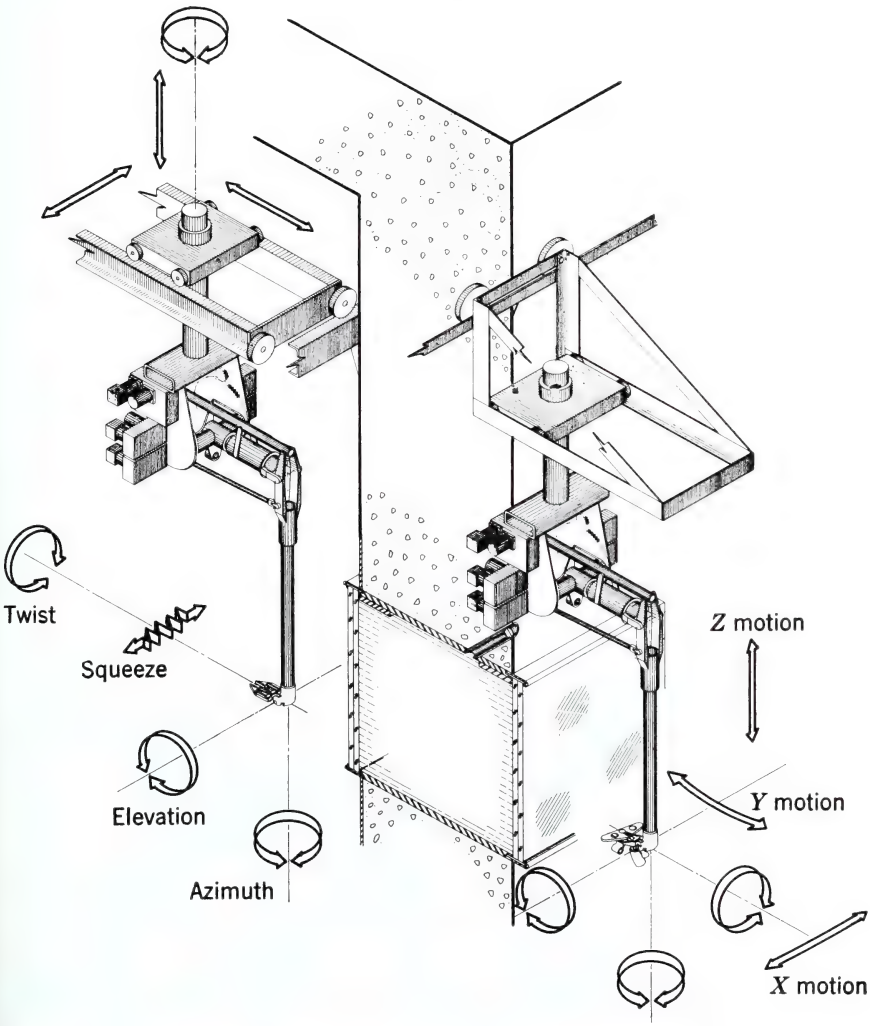


FIG. 27-10 *Model 2 master-slave electric manipulator.*

the gearing was such that the maximum speed in the X, Y, and Z directions was 24 in./sec. It was then found that most of the operators wanted to exceed this speed sometimes. Since the servomotors are of the a-c type, they strongly resist going faster than this maximum speed. When the

operator felt this restriction of speed, he tended to push harder. This, of course, makes meaningless the whole idea of force reflection. The kinesthetic senses in the muscles tend to become saturated, and the operator can no longer be sensitive to collision or other loads reflected from the slave arm. Because of this human factor, the maximum speed for the X, Y, and Z motions was increased to 36 in./sec. This speed satisfied all the operators tested. However, this experience should not be used to indicate that manipulators should always have a speed of about 36 in./sec. It should be possible to have a lower maximum speed if some type of warning device were incorporated that signaled the operator when he was nearing maximum speed. This idea has not yet been tested.

The master-slave electric manipulator, model 3, has been built, tested, and put into operation. It has a load capacity of 50 lb in any direction. Figure 27-11 is a line drawing of this manipulator, with the master and

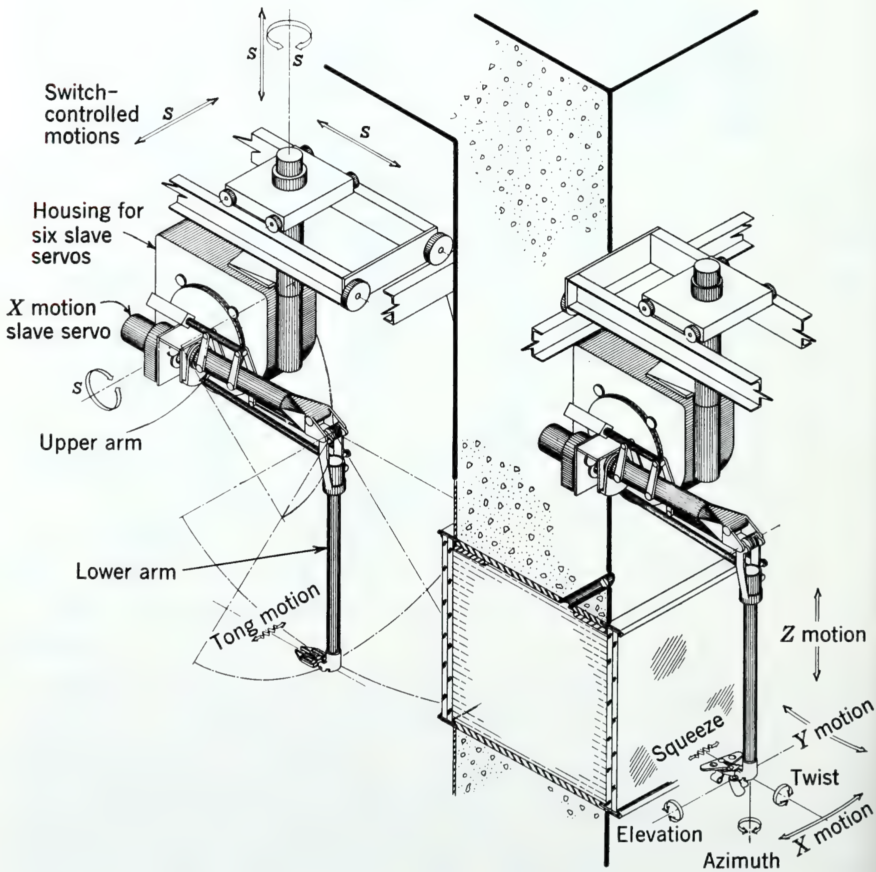


FIG. 27-11. *Master-slave electric manipulator, model 3.*

slave arms mounted on independent overhead bridge and rail systems as they are at present in use.

Figure 27-12 shows a pair of these manipulators mounted on an electrically powered, wheeled vehicle. This device is called a slave-robot and is capable of traveling over an ordinary floor.

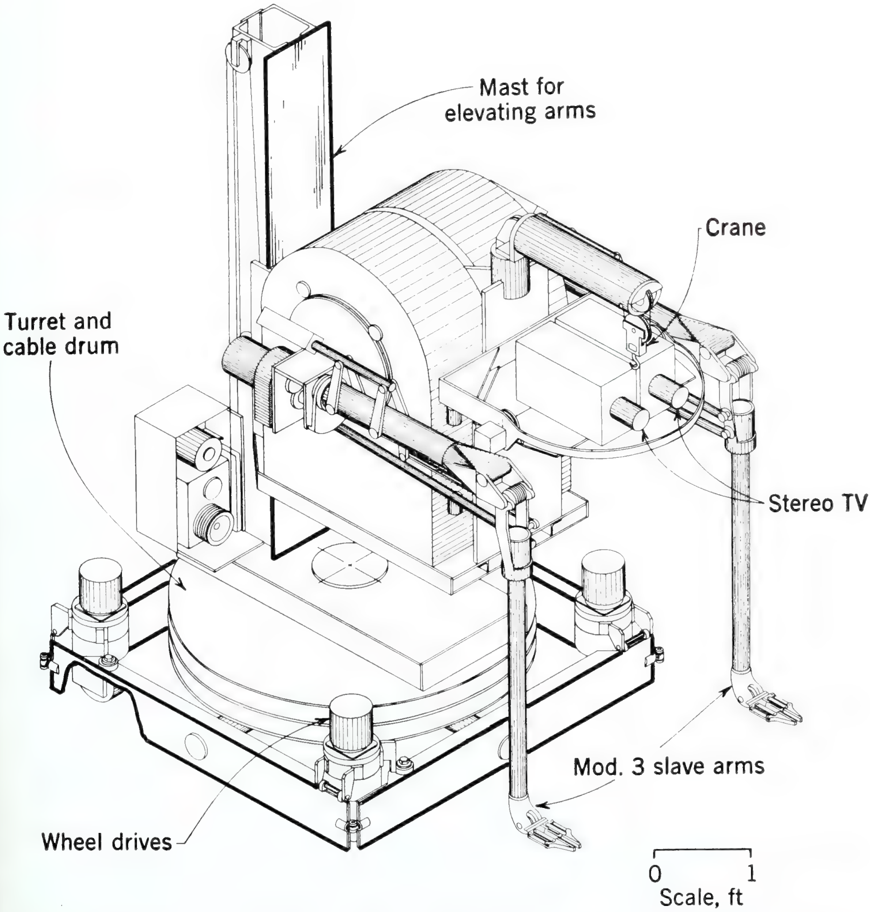


FIG. 27-12 *Slave-robot equipped with two master-slave electric manipulators, model 3.*

The force-reflecting servos for this manipulator are modifications of those shown in Figures 27-8 and 27-9. Four servomotors are geared to the slave-servo assembly, and two servomotors are geared to the master-servo assembly. The gear ratio of the master-servo assembly is only two-thirds that of the slave-servo assembly. The rest of the circuitry is similar to that shown in Figure 27-8 except that switches are provided

that change the input-output force ratio from 1:1 to 3:1; the squeeze-force ratio can also be 9:1. The maximum load capacity of the manipulator when it is operating in the 1:1 force reflection is about 17 lb. When it is operating in the 3:1 force ratio, the maximum load is 50 lb. The master arm never reflects more than about 17 lb. This magnitude has been selected for safeguarding the human operator from abuse in case of electronics malfunction. The smaller number of motors and the lower gear ratio also reduce the effective mass felt by the operator. With the handles used, it would be difficult for the operator to handle more than about 17 lb. It is also probable that the kinesthetic-force indications are less sensitive at the higher loads than the lower loads; that is, the kinesthetic sensation may be about as good with only one-third the force being reflected when the output loads are high as with full load-force reflection. However, this has not been systematically measured.

The effective inertia of this manipulator is 15 lb for the X and Y directions and 26 lb for the Z direction. These are the apparent masses that the operator feels when operating the manipulator with the force ratio set at 1:1. The mass in the Z motion is considerably higher than the corresponding mass for the mechanically connected master-slave manipulator, model 8, and causes noticeable increased awkwardness. The angular inertia of the elevation motion is 178 lb-in.^{2}, or 4.95 lb at the center of the tongs, the inertia of the twist motion is 28 lb-in.^{2}, and the inertia of azimuth is 402 lb-in.^{2}, or 7.7 lb at the center of the tongs. The values are the apparent inertias felt by the operator when the force ratio of the manipulator is 1:1. These inertias are considerably higher than the corresponding inertias for the model 8 manipulator, and there is considerable reduction in the ability of an operator to feel the load. When the force ratio is 3:1, these apparent inertias are considerably reduced and the performance of the manipulator seems to be improved. Since the 3:1 ratio has worked out quite satisfactorily, there seems to be no reason why the force ratio could not be increased for manipulators having higher load capacities.

This manipulator incorporates electrically operated brakes in each of the slave-servo gearboxes. The operator can apply these brakes by actuating a conveniently located switch on the master handle. These brakes are useful for locking the manipulator whenever the operator wants to hold an object in a certain position. However, the primary purpose of the brakes is that of automatically locking the slave arm in case of power failure or malfunction of the servo system.

Both the mechanical master-slave manipulator, model 8, and the master-slave electric manipulator, model 3, can perform work ten or more times as fast as a unilaterally controlled manipulator such as the one shown in Figure 27-3. This high speed of operation is due to both the bilateral drives and controls and the natural movement of the slave arm

as it follows the master arm and handle. A bilateral control system, with low inertia and friction, is extremely useful for manipulators that must perform complex and delicate operations.

Even though the master-slave manipulator is reasonably dexterous, it is far inferior to the human arm and hand. Therefore, some of the following features should be designed into future manipulators: better wrist joints, multiple-fingered hands, better force reflection, and tactical force reflection. Because of the extensive possibilities of improving manipulators, it is believed that these devices are in the infancy stage of their development.

The general-purpose manipulators described here are controlled completely by a human operator. It is possible, of course, to program the motions and forces that the manipulator should follow, provided that these motions and forces are known beforehand. It is also possible to program a general-purpose manipulator arm for repeated operations. Some arms have been developed specifically for this purpose, but these units do not have full general-purpose capabilities. Usually they do not incorporate many of the functions that the human being exercises when he performs similar tasks, and therefore the range of work that these devices can handle is limited.

Chapter 28

EXPLORING THE INNER SPACE OF LIVING CELLS BY MICROSURGERY

M. J. Kopac[*]

THE EXPLORATION BY MICROSURGERY OF THE VARIOUS STRUCTURES, OR INNER space, of living cells presents many intriguing problems. Most cells are so small that they must be magnified 100 to 2,000 times before the more obvious structures such as nuclei, nucleoli, chromosomes, or mitochondria can be properly seen. Because cells are delicate structures, they must be handled with extreme care. Furthermore, living cells are constantly changing, and so the structures and dynamics of a cell in one instant may be strikingly different in another instant.

Obviously, if one is to perform microsurgery on cells, the surgical instruments, consisting of fine glass microneedles or micropipettes, should be much smaller than the cellular structures, and one must insert these surgical instruments into the cell with great precision and delicacy in order to avoid excessive injury to the cell [1]. Advances in this field of research have been made possible through the development of complex mechanical devices by which microsurgical instruments are placed into cells [5]. With these highly precise mechanical devices, known as micromanipulators, one can insert needles or pipettes into a cell with delicacy so precise that any selected part of the cell or a nucleus can be removed and implanted into another cell or nucleus. Procedures and instrumentation for handling and measuring small volumes of liquid have also been developed. Microinjectors have been designed and constructed that can measure or control volumes of less than 1 $\mu\mu$l.

Thus, through optical and electronic devices coupled with suitable

[*] All-University Department of Biology, New York University, New York, N.Y.

The work discussed in this chapter was supported in part by Grant C-2018C7 from the National Cancer Institute, Public Health Service, Bethesda, Md., and by Grant DRG-271B from the Damon Runyon Memorial Fund for Cancer Research, Inc., New York, N.Y.

The volumetric submicromanipulators, the differential piston microinjectors, and the multimicromanipulator-microinjector system were built and the servo-controlled micromanipulators are being built by Jack Harris, whose skills as an instrument designer and builder are gratefully acknowledged.

micrurgical instrumentation, the manual skill of the micrurgist can be extended to the microsurgery of cells or cell structures. Facilities are now available for positioning the tip of a microneedle or micropipette within the resolution limits of the light microscope. Various parts of cells such as nuclei, nucleoli, or chromosomes, can be transplanted from one cell to the cytoplasm or nucleus of another cell [4, 6].

MULTIMICROMANIPULATOR-MICROINJECTOR SYSTEMS

An example of a multimicromanipulator-microinjector system is illustrated in Figure 28-1. This is highly sophisticated instrumentation designed for solving difficult micrurgical problems. Specifically, one problem is to produce microsurgical translocations in chromosomes and then to transplant the modified chromosomes into the nucleus of another cell. This instrument consists of six micromanipulators plus various types of microinjectors capable of mounting 10 microneedles or micropipettes with no interference between any of them or the micromanipulators.

For the microsurgery of chromosomes, the donor cell must be held in place by microelastimetry. Then the chromosome is cut at the selected position. With one of the micropipettes, controlled by the submicrovolumetric unit, shown in greater detail in Figure 28-2, the cut fragment is brought in contact with the end of another chromosome, whose telomere has been made sticky by the local application of radiomimetic drugs [4]. After the cut fragment has fused with the chromosome, the entire unit is transplanted into a new cell by means of the second volumetric submicromanipulator.

Two of the six micromanipulators, mounted in the rear position, are ball-bearing slide micromanipulators that carry large micropipettes and are used primarily for holding cells by microelastimetry [10]. The pressure control of the fluid in the micropipettes, so that the necessary suction for holding cells is produced, is controlled by the Harvard Apparatus Withdrawal-Perfusion pump (model 600-900). This pump is set into operation by a foot switch. A two-way valve joins the pump to either one of two micropipettes.

These micromanipulators, adapted for holding and positioning cells, are capable of being moved with a high degree of precision. Ball-bearing slides are essentially friction-free. Backlash is eliminated by pressing the slide against the feed screw with a spring-loaded plunger. Large dials, 5 cm in diameter and graduated into 100 units, serve as feed-screw handles. Since the pitch of the feed screw is 0.5 mm, each division corresponds to a slide displacement of 5 μ. The rotation of the dials may be easily controlled to one-fifth division, a slide displacement of 1 μ.

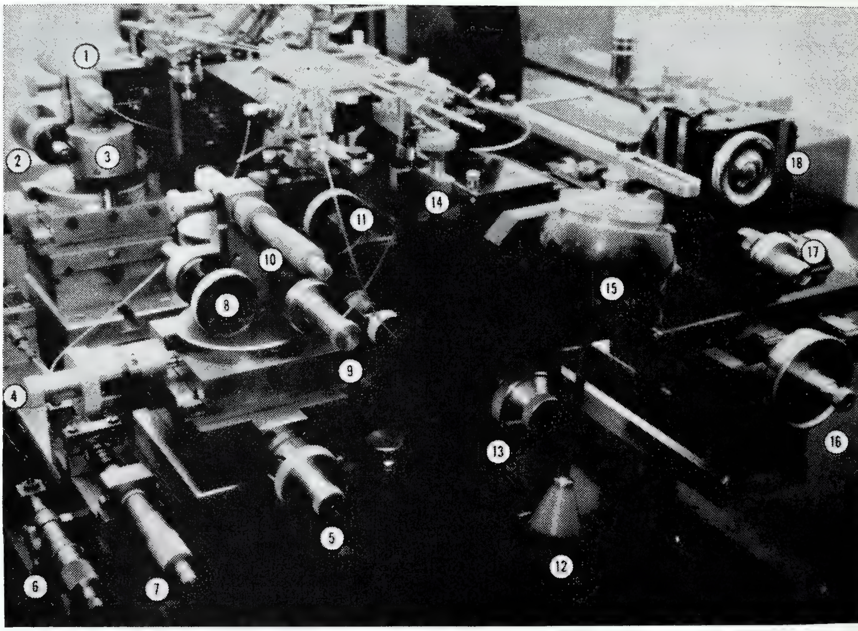
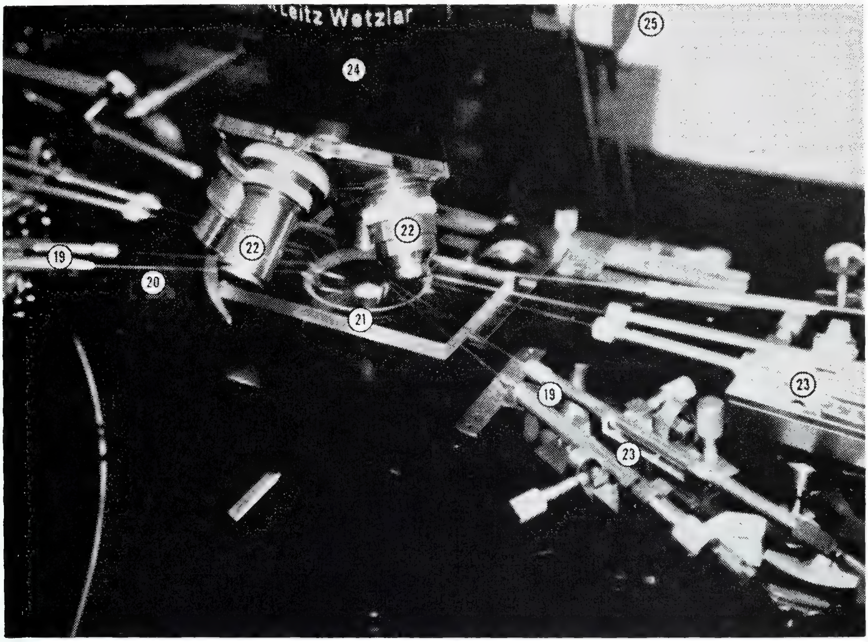


FIG. 28-1 *Multimicromanipulator-microinjector system. (a) This instrument consists of six micromanipulators, of which three are seen in the foreground. A similar set is mounted to the left of the microscope. The volumetric submicromanipulator is seen on the left-foreground side. The micromanipulator in the middle is the 1959 Leitz lever-controlled unit. The unit on the right is the precision ball-bearing slide micromanipulator. This instrument has a single-needle holding clamp, whereas the others have the Leitz double-needle holding clamps. Just below the volumetric submicromanipulators are mounted two microinjectors. Altogether, six micropipettes can be controlled by the volumetric submicromanipulators, the microinjectors, and the withdrawal-perfusion pump. Only the Leitz units carry microneedles.*

1. *Pillar assembly for supporting volumetric controls*
2. *Fine volumetric control on volumetric submicromanipulator*
3. *Housing for worm-gear feed-screw drive for micropiston*
4. *Micrometer head for horizontal movement on left volumetric submicromanipulator*
5. *Micrometer head for horizontal movement*
6. *Fine volumetric control on accessory microinjector*
7. *Coarse volumetric control*
8. *Fine volumetric control on right volumetric submicromanipulator*
9. *Fine vertical movement*
10. *Coarse volumetric control*
11. *Coarse vertical movement*
12. *Lever for horizontal movements on Leitz (1959 model) micromanipulator*
13. *Coarse-fine controls for vertical movement*
14. *Coarse control for horizontal movement*
15. *Ball-sphere segment for fine horizontal movements*
16. *Control knob for horizontal movement on precision micromanipulator*
17. *Fine vertical movement*
18. *Coarse vertical movement*



(b) Close-up photograph with 6 micropipettes and 4 microneedles. The operating chamber is a Romicron depression slide with a flat, polished bottom. Long-working-distance objectives permit operation of microneedles and micropipettes between object and objective. With a circular operating chamber, accessible from any position, the 10 needles and pipettes may be grouped for performing almost any microsurgical procedure in an efficient manner.

19. *Micropipette holders*
20. *Glass micropipettes*
21. *Romicron glass slide serving as microsurgical chamber*
22. *Microscope objectives*
23. *Double-micropipette-holder clamp*
24. *Microscope-body tube*
25. *Fine adjustment knob on microscope*

The initial friction of the slide is not appreciably greater than the friction developed after the slide is in motion, and the slides, therefore, instantly respond to the slightest rotation of the feed-screw dials. Modified coarse-fine microscope focusing mechanisms provide the vertical movements.

The two instruments shown in Figure 28-1*a* were built in 1937. They have been used in conjunction with the volumetric submicromanipulators [2] and for microsurgery of chromosomes [4]. Despite the age of the two units, they still are capable of exquisitely precise movements. Neither the ball-bearing slides nor the feed screws have required any adjustment or compensation for wear. The main advantage resulting from the adaptation of ball-bearing slides to micromanipulators is that, by spring

loading, backlash is neutralized. Response is instantaneous and delicate, the delicacy being dependent solely on the actuator, whether controlled by feed screws or lever.

The Leitz 1959 lever-controlled micromanipulators, equipped with double-needle holders, are mounted in the middle as shown in Figure 28-1*a*; one of these micromanipulators is shown in greater detail in Figure 28-3. These micromanipulators carry delicate glass microneedles used primarily for cutting chromosomes. By crossing the opposing microneedles near their tips and then moving them apart, one has essentially the same facility possible with a pair of wire cutters. Chromosomes can be cleanly and precisely cut at any desired position.

These lever-controlled micromanipulators are constructed with the lever mechanically coupled to ball-bearing horizontal slides [9]. Coarse and fine vertical movements are produced by rotating two coaxial knobs. The model illustrated in Figure 28-3 was available from 1956 to 1960. A more massive unit, capable of carrying heavy loads such as a double-needle holder, was produced in 1961. The double-needle holder has a ball-and-socket clamp that permits orientation and preliminary positioning of the microinstruments. The ratio of lever to slide movements, ranging from 16:1 to more than 800:1, can be regulated by adjusting the position of the ball segment. Coarse adjustments of considerable range are possible for all three movements. In addition, the horizontal movements can be inclined at an angle of 15° from the horizontal plane by a rack and pinion.

Two volumetric submicromanipulators, equipped with Leitz double-needle holders so that two micropipettes can be simultaneously controlled by each instrument, are mounted in front as shown in Figure 28-1*b*; one is shown in greater detail in Figure 28-2. Volumes of the order of a micromicroliter can be handled efficiently with the volumetric submicromanipulator. One micropipette is controlled with the volumetric controls provided by the instrument itself. The second micropipette is connected either to one of the microinjectors used for microelastimetry or to the differential-piston microinjector designed for subcellular transplantation.

One of the volumetric submicromanipulators has a piston diameter of 0.005 in. and the other a diameter of 0.0158 in. The steel pistons are Class A cylindrical standards made to a tolerance of ± 0.00001 in. The pistons operate in lapped steel barrels, in which the length of contact between the piston and barrel is six to eight times the diameter of the piston. The movements of these fine pistons are regulated by carefully constructed mechanisms, each consisting of a feed screw and a worm gear equivalent to a screw with 125,000 threads per inch. One full turn of the volumetric control knob moves the piston 8 μin. The linear dis-

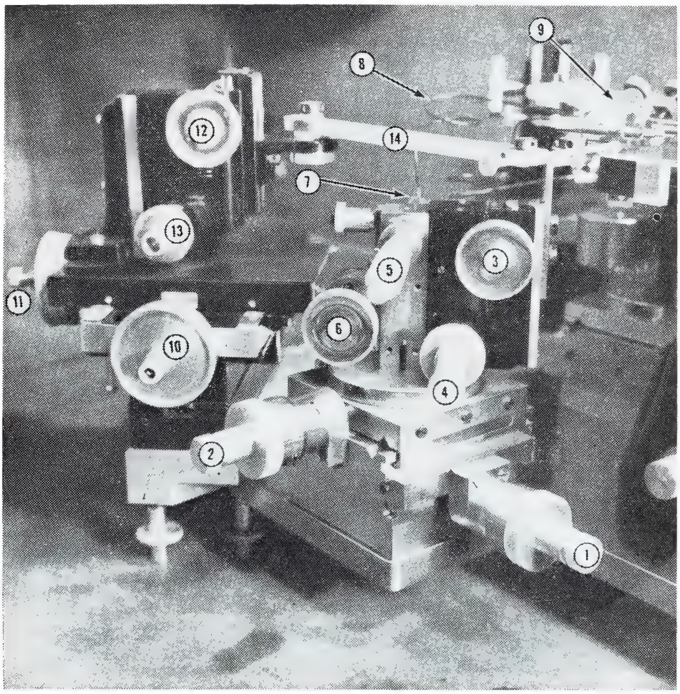


FIG. 28-2 *One of the two polished metal volumetric submicromanipulators (shown as No. 8, 9, and 10 of Figure 28-1) and one of the two black metal precision micromanipulators (shown as No. 16, 17, and 18 of Figure 28-1).*

1. *Micrometer head for controlling horizontal movement on volumetric submicromanipulator*
2. *Micrometer head for controlling horizontal movement*
3. *Coarse control for vertical movement*
4. *Fine control for vertical movement*
5. *Coarse volumetric control*
6. *Fine volumetric control*
7. *Volumetric chamber*
8. *Tubing from volumetric chamber to micropipette holder*
9. *Micropipette holders and clamps*
10. *Control knob for horizontal movement on precision micromanipulator*
11. *Control knob for horizontal movement*
12. *Coarse control for vertical movement*
13. *Fine control for vertical movement*
14. *Arm for supporting micropipette-holder clamp*

placement of the piston is measured with an electronic micrometer consisting of an RCA 5734 mechanoelectronic transducer tube with one stage of d-c amplification. Linear displacements as small as 0.1 μin. can be measured. The smaller piston permits volume measurements down to 0.03 $\mu\mu$l. Samples of cytoplasm ranging from 0.3 to 10 $\mu\mu$l can be measured,

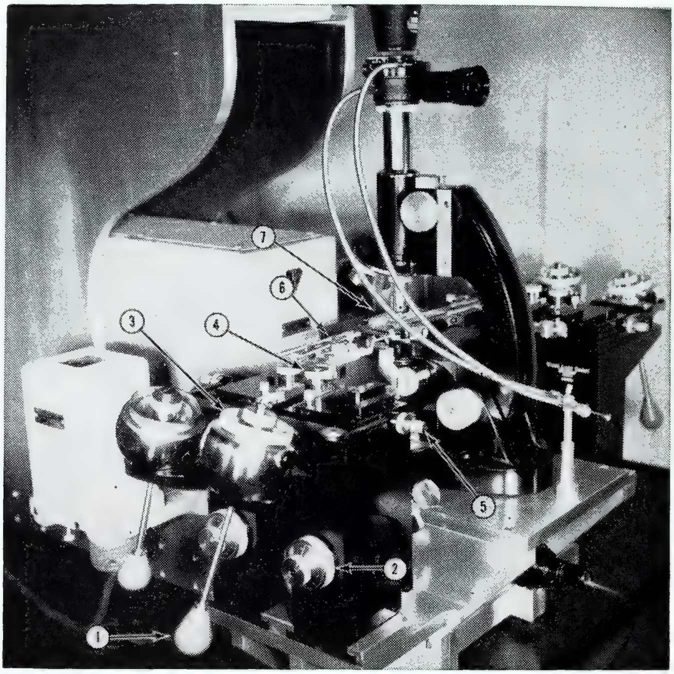


FIG. 28-3 *One of the two Leitz micromanipulators (shown in Figure 28-1).*
1. *Lever control for fine horizontal movements on Leitz (1959 model) micromanipulator*
2. *Coarse-fine controls for vertical movements*
3. *Ball-sphere segment for horizontal movements*
4. *Coarse control for horizontal movement*
5. *Coarse control for horizontal movement*
6. *Micropipette holders and clamps*
7. *Microsurgical chamber*

removed, and transplanted into other cells. The larger piston, which gives ten times the volume of the smaller piston for the same linear displacement, is especially useful for transplanting nucleoli or chromosomes.

The original intention in planning the volumetric submicromanipulators was to integrate the microinjector or submicrovolume control with the micropositioning of micropipettes. Ball-bearing slides were used to provide horizontal movements and spring loading to avoid backlash. Motion is imparted by metric screws equipped with large micrometer heads. These micrometers provide a fine screw with a pitch of 0.5 mm and an adjustable nut that can compensate for wear. Each division corresponds to 2.5 μ of travel of the feed screw and, accordingly, of the horizontal slide. The combination of micrometer-driven, ball-bearing, and spring-loaded slides produces an extremely smooth movement.

A coarse-fine microscope focusing mechanism was modified for the vertical movement. The coarse adjustment permits quick raising or lowering of the micropipette; the fine adjustment provides a delicate movement for the proper vertical placement of a micropipette.

Although the volumetric submicromanipulators have been a most useful adjunct for accomplishing difficult microinjections or subcellular transplantations, they do have serious drawbacks. Since the control of volume is provided by fine pistons and is actuated by complex, carefully fitted worm-gear feed-screw mechanisms, the production of such instruments is expensive and requires the skill of an accomplished instrument maker. The more difficult problem with fine pistons is that of constructing packing devices that will not leak. A tight packing cannot be used, for there is always the danger that an ultrafine piston will freeze and, consequently, will bend or buckle instead of passing through the packing into the volume chamber. Recently, this difficulty was solved by adapting the principle of the differential piston [3]. Two pistons, approximately 1 mm in diameter, are mounted on a carrier bracket. The pistons enter a volume chamber from opposite ends. As one piston moves into the chamber, the other piston moves out. One piston is slightly larger than the other so that a volume differential results. The smaller the difference in diameter of the two pistons, the smaller the volume differential per unit of linear travel. Since the pistons are relatively large and of hardened steel, a Teflon packing gasket can be used and thus any danger of leaking or buckling of the piston is prevented.

The two microinjectors mounted in front of the volumetric submicromanipulators in Figure 28-1*a* have been in service for many years. These microinjectors proved to be especially useful for the study of surface chemical properties of cytoplasmic proteins at oil-water interfaces by the drop-retraction technique [8]. The same units were subsequently used for preparing enzyme-substrate reaction droplets, ranging from 0.0005 μl to several microliters, as required in microdilatometry [2, 7]. More recently, the same units have served for holding cells by microelastimetry [3]. Micropipettes are filled by the coarse volume control, consisting of a Luer syringe and a micrometer head. The fine control, made from a modified micrometer, is appropriate for managing smaller volumes.

SERVO-CONTROLLED MICROMANIPULATORS

There is no question that the micromanipulators of the future must be partly or completely automated. Automation will (1) provide greater ease in operating multimicromanipulator and microinjector systems and (2) permit the use of massive micromanipulators containing ball-bearing

slides and driven by feed screws, which give the greatest precision. Through servomotor drives, the micrurgist will be able to activate simultaneously any number of micromanipulators either with push-button controls or semiautomatically through programming information stored in punched or magnetic tapes.

The servomotor-driven multimicromanipulator system now being constructed has the following general specifications: There are four micromanipulators, each consisting of three-way ball-bearing slides of the same type as used in the volumetric submicromanipulator, only larger. There are two methods for driving both the horizontal and the vertical slides, manual control and servomotor drives.

The manual controls are massive micrometer heads with the 1-in. spindle (0.25 in. OD) directed into a volume chamber. On each of the spring-loaded slides there is a larger piston (0.5 in. OD) emerging from a volume chamber built onto the slide, the receiving piston chamber. Fluid displaced by the micrometer head is transmitted to the slide or receiving piston chamber via a flexible metal tube. The piston diameters have a ratio of 2:1 so that 4 mm of travel of the micrometer spindle produces 1 mm travel on the slide.

The hydraulic transmission of the servo-driven piston or manual piston provides the smoothest possible method of actuating the ball-bearing slides. Additional reduction of friction between the drives and slides is achieved by increasing the diameter ratios of the driving and the receiving pistons.

The servo controls for each micromanipulator consist of three Kearfott R112-2B servomotors, each geared down to a feed screw that displaces a piston (0.125 in. OD) in a volume chamber. This chamber is connected to the receiving piston chamber on each slide via a flexible metal tube. A 20,000-ohm precision potentiometer is connected to the gear train for indicating the position of the slide.

The general relations of the various operating components associated with the micromanipulators per se are shown in the schematic diagram (Figure 28-4). The various controls, consisting of reference potentiometers, relays, and push buttons, are mounted on a separate console. Two reference potentiometers are used for each slide and are adjusted so that with one the micropipette will be moved on target and with the other it will be moved off target.

Switching from the on-target to the off-target position, or vice versa, is accomplished by means of a push-button relay circuit. If the micropipette is in the off-target position and the on-target potentiometer is switched in, there will be an error between the indicating potentiometer on the servo drive and the reference potentiometer. The servomotor, through the servo amplifier, will correct this error by moving the micro-

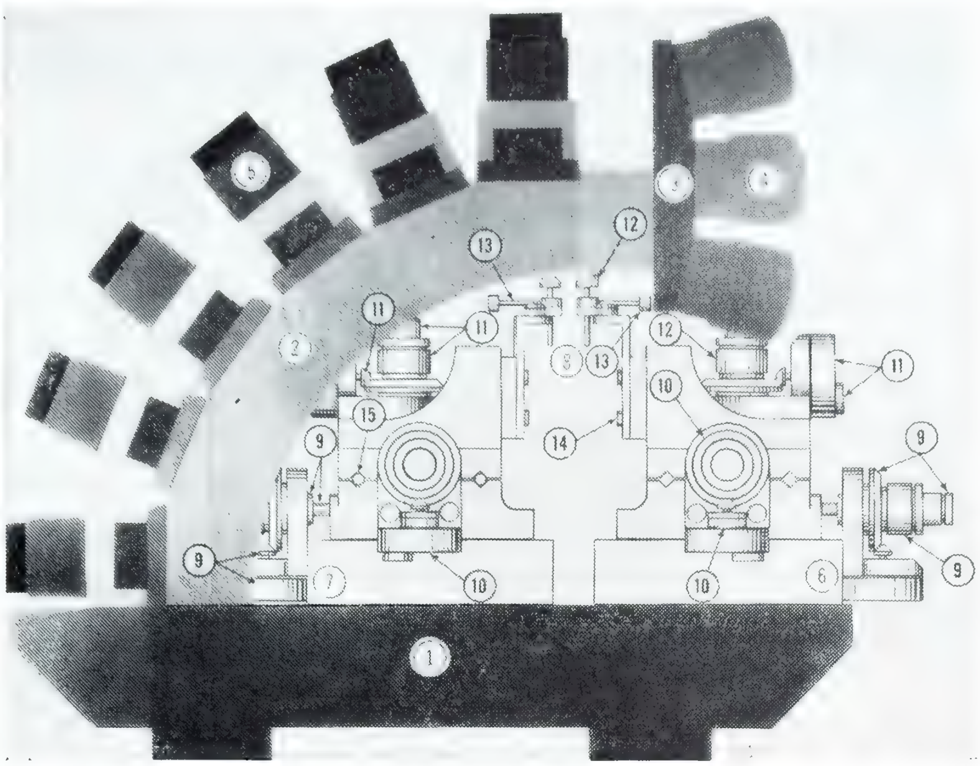


FIG. 28-4 *Schematic diagram of servo-controlled micromanipulator.*

1. *Base*
2. *Arm for supporting six servomotor drives used on left side*
3. *Mounting for manual controls consisting of oversize precision-micrometer heads*
4. *Three micrometers for each micromanipulator, one for each movement*
5. *One servomotor drive, of six mounted on left side, each consisting of a servomotor, gear train, position-indicating potentiometer, and terminal strip*
6. *Front and*
7. *Rear micromanipulators on left side, with similar pair on the right side*
8. *Mounting brackets and clamps for micropipette holders*
9. *Details of piston-driving mechanism for horizontal movement (Y axis)*
10. *Details of piston-driving mechanism for horizontal movement (X axis)*
11. *Details of piston-driving mechanism for vertical movement (Z axis)*
12. *Micropipette-holder clamp*
13. *Lever screw for micropipette-holder clamp*
14. *Screw for preliminary vertical adjustment of micropipette-holder clamp*
15. *V-way for ball-bearing slide*

pipette and the indicating potentiometer to the correct position—in this instance, the on-target position. Thus, 12 push buttons can control all three movements on each of the four micromanipulators. At any time the reference potentiometers can be changed to provide different on-target or off-target positions for the micropipettes. With this arrangement, any one of the four micropipettes will always be either in the right place or clearly out of the way.

Each micromanipulator is furnished with at least one differential-piston microinjector, with the actuator driven by a servomotor. This unit can be used for both microinjection and microaspiration. Most

important, these microinjectors can be programmed for performing aspirations at different rates, as required for dislodging and removing certain subcellular structures.

A microscope stand has been designed, and a prototype now being built is to be integrated with the servo-controlled micromanipulators. The microscope has an inverted-optics arrangement using a modified Leitz binocular-phototube attachment that provides a satisfactory means of combining video-scanning techniques with microscopy. In the design of this microscope, the mechanical stage was redesigned to incorporate both conventional coarse movements and fine adjustments that can be controlled by servomotor drives. The fine adjustment of the microscope will also be driven by a servomotor so that all operations involving the microscope as well as the micromanipulators and microinjectors can be remotely and automatically controlled through the use of punched- or magnetic-tape programming.

With the micromanipulators, the mechanical stage, and the fine adjustment of the microscope under servo control, it will be possible to utilize some of the techniques employed in tracking missiles. Video-scanning facilities will provide the information needed for tracking. As one can track an Atlas missile photographically, it will be possible to track an active amoeba under the microscope and to place microneedles and micropipettes into the proper operating positions within the amoeba by automatic means.

Undoubtedly, future explorations of living cells by microsurgery will be in the field of experimental cytogenetics. Nuclei, chromosomes, or DNA (deoxyribonucleic acid) can be implanted or injected into cells. Although nuclei are brought together from two different cells by normal fertilization mechanisms, microsurgery makes possible the bringing together of nuclei that normally would not be able to accomplish this step spontaneously. Moreover, selected chromosomes from other cells can be transplanted into a cell in mitosis so that these chromosomes may merge with the chromosomes already present. Highly refined DNA may be similarly added to a cell before, during, or after mitosis. Perhaps the utilization of subcellular transplantation techniques will help solve some of the problems dealing with cytoplasmic inheritance. Without proper instrumentation, however, such experimentation on living cells will not succeed.

BIBLIOGRAPHY

1. Chamber, R., and M. J. Kopac: Micrurgical Technique for the Study of Cellular Phenomena, in R. M. Jones (ed.), "McClungs's Handbook of

Microscopical Technique," 3d ed., pp. 492–543, Paul B. Hoeber, Inc., New York, 1950.

2. Kopac, M. J.: Cytochemical Micrurgy, in G. H. Bourne and J. F. Danielli (eds.), "International Review of Cytology," vol. 4, pp. 1–29, Academic Press Inc., New York, 1955.

3. ———: Experimental Studies on Malignant Nucleoli, in "Cell Physiology of Neoplasia," pp. 97–128, University of Texas Press, Austin, Tex., 1960.

4. ———: Exploring Living Cells by Microsurgery, *Trans. N.Y. Acad. Sci.*, vol. 23, pp. 200–214, January, 1961.

5. ———: Micromanipulators: Principles of Design and Operation, in W. Nastuk (ed.), "Physical Techniques in Biological Research," vol. 6, Academic Press Inc., New York, in press.

6. ———: Micrurgical Studies on Living Cells, in J. Brachet and A. Mirsky (eds.), "The Cell," vol. 1, pp. 161–191, Academic Press Inc., New York, 1959.

7. ———: Submicro Methods in Enzymatic Cytochemistry, *Trans. N.Y. Acad. Sci.*, vol. 15, pp. 290–297, June, 1953.

8. ———: The Surface Chemical Properties of Cytoplasmic Proteins, *Ann. N.Y. Acad. Sci.*, vol. 50, pp. 870–909, February, 1950.

9. Seidel, P.: Bauweise und Verwendung von Mikromanipulatoren, in H. Freund (ed.), "Handbuch der Mikroskopie in der Technik," vol. 1, pt. 1, pp. 591–613, Umschau Verlag, Frankfort, 1957.

10. Tyler, A., et al.: Membrane Potential and Resistance of the Starfish Egg before and after Fertilization, *Biol. Bull.*, vol. 111, pp. 153–177, August, 1956.

Part G

HUMAN FACTORS IN THE MECHANIZATION OF HUMAN FUNCTIONS

Chapter 29

SYSTEMS THAT LEARN

*L. O. Gilstrap** *and R. J. Lee**

IT IS IMPORTANT TO KNOW HOW FAR MACHINES CAN GO IN REPLACING OR supplementing the human operator. In general, the degree of replacement or supplementation is dependent upon the ability of the machine to utilize raw data, as well as its ability to solve problems without being told how. A system, whether it is a man or a machine, can be completely autonomous in doing either of these only to the degree that it has learning ability. The meaning of data, particularly of new symbols, must often be learned by association, whereas often it is only by experimentation, that is, by trial-and-error learning, that the solution to a particular problem can be found. Machines able to experiment and to learn are said to exhibit artificial intelligence, particularly if their learning capability is fairly general.

The processes that such automata use can be categorized as some combination of sensing, pattern recognition or perception, decision making, logical induction, deduction, and analog-function formation, control, and display. Their learning is accomplished in terms of some criteria supplied either by a human trainer or by a goal that is made a part of the machine. The goal may be general or specific and may itself be altered in some respects by learning. Thus, the learning process for the machine is a matter of becoming more proficient in the meeting of the criteria supplied by the goal or the trainer.

Three types of learning machines can be found described in the literature, (1) machines made up of networks of artificial neurons or other esoteric building blocks, (2) machines that are basically digital computers but that are programmed heuristically, and (3) machines that have been designed to illustrate a specific learning principle or mechanism. Although the earliest reference to a learning automaton was made in 1913 [16, 49],[1] the field was apparently dormant until the 1940s, when Rashevsky [42], Householder and Landahl [21], McCulloch and Pitts [30], Wiener [59], and Hebb [18] published various pertinent papers and books. In the 1950s important concepts were brought forth

* Adaptronics, Inc., Annandale, Va.

[1] Because Chapters 29 and 30 are closely related, the references for both chapters are presented at the end of Chapter 30.

by Walter [57, 58], Ashby [1], Farley and Clark [15], Lee [27, 28], Uttley [54], Rosenblatt [47], Newell, Shaw, and Simon [37], and numerous other researchers.

The general growth of interest in this field has been led by the increasing demands for data processing and problem solving in military operations under unknown and varying environmental conditions. As a result, machines that require neither the special coding of input data nor the complete specification of problem-solving instructions must be devised. Such machines must be capable of relatively autonomous functioning in natural environments. With devices possessing this increased capability, however, the more specific functional requirements must be carefully examined. Some of the functional requirements for these machines are summarized in this chapter. A number of different learning automata are described and compared in Chapter 30.

FUNCTIONAL REQUIREMENTS FOR SYSTEMS THAT LEARN

In examining a problem area for the possible applicability of learning automata it is essential to know those items of performance which are attributes of the learning ability as well as those which are attributes of the particular approach or techniques used. If the ability to achieve desired performance in a varying environment is taken as a prime requirement for a learning machine, it is apparent that such a machine must be able to construct within itself a model of the environment that can be used by the machine to obtain either responses or response requirements from a given stimulus. Such a model of the environment need not be a literal, explicit model of the specific objects in the environment, but it must be accurate as a model of the patterns of events found in the environment. The problem of the kind of events or entities that can be represented within an automaton has been treated by Kleene [25] and MacKay [31] and will not be examined here. It is sufficient to point out that most events of interest can be represented within an automaton. This includes events that are conventionally represented by number, magnitude, or class, and these are, in general, adequate for most applications. How any system capable of learning can construct within itself a capability for achieving a desired performance in a varying environment has been conjectured by Ashby [1] to consist of a process similar to natural selection. An automaton or other type of system can make internal changes, comparable with mutation, the new functions then being selected or retained depending upon their value in meeting the overall system criteria. The idea of using a natural selection of circuits in a machine, based upon a random search for a suitable logical

function and survival of useful functions, has been discussed by Lee [27] in terms of providing the machine with the ability to learn, to make and test hypotheses, and to display initiative, inductive reasoning, and creativity.

More recently Campbell [12] has discussed how an automaton or living system can gain a knowledge of its environment. He has suggested that it does so by means of two fundamental processes, blind variation, i.e., random or pseudo-random experimentation by the automaton, and selective survival of the useful results from such experimentation. This would be equivalent to the construction of a model of the environment by natural selection of functions within the system. The correctness of Campbell's suggestion may be debatable, but the elements of blind variation and selective survival do appear in many learning machines that have been proposed to date. Even in the heuristic programs that enable the computer to learn such diverse activities as chess playing [35], theorem proving, and musical composition, the rules of thumb employed in these programs are derived from observations of human beings engaged in similar activities, and it is evident that such rules of thumb are often developed by trial-and-error methods. In any event, the blind-variation and selective-survival hypothesis is useful in that most of the functional requirements of learning automata can be derived from this hypothesis:

1. Blindness (randomness or pseudo randomness) is necessary in the experimental procedure to avoid such phenomena as nonadaptive repetition in behavior and becoming trapped at local maxima rather than finding absolute maxima in devices that search surfaces in function space[2] or in physical space.

2. Spontaneity is necessary in the blind-variation or experimentation process. Not all learning systems exhibit this spontaneity, and those which do not can learn only from the stimuli presented to them. It is clear that the amount of learning can be increased by allowing the system to initiate a learning situation.

3. The automaton must be directed by a set of criteria whereby the useful variations in behavior can be distinguished from the nonuseful. A set of criteria is equivalent to a goal or set of goals; the goal, rather than the criteria, is defined first in some systems. In some learning systems the criteria are provided by a human trainer.

[2] Function space is a mathematical abstraction. It may have as its coordinates selected variables of the system, such as temperature, light intensity, and kinematic variables, or it may be defined as having coordinates that are the internal variables, or states, of the machine. Searching of function spaces constitutes trying different values of these variables. The two types of function spaces are related by a mapping that, in the more sophisticated cases, is many to many.

4. For problem solving and learning in unknown environments an automaton must be capable of at least as much variety in its internal behavior as there is in the portion of the environment that it needs to control or model. Since the complexity of an environment, as viewed by the automaton, is dependent upon the demands posed by the automaton's goals, both the goal and the environment must be considered in determining the requisite complexity of an automaton.

5. From the blind-variation process, it is further evident that an automaton must have many degrees of freedom in its overt behavior. Not only must the learning portion of the automaton have sufficient internal complexity to match the complexity of its environment, but it must also be able to translate this complexity into action.

6. For the more general automaton, step-by-step learning is highly desirable in order to reduce the learning or adaptation time. This implies the ability to learn a new form of behavior while retaining the old form. This ability is a characteristic of systems that Ashby [1] has defined as multistable. It is largely that of being able to learn by experience and may also imply the need for a special memory system.

Taken together, these are the requirements of an intelligent machine. Not all investigators would agree that all these requirements are essential or that all requirements are included, but the list does represent a synthesis of the more commonly found requirements. From the behavioral point of view, these are the requirements for machines capable of exhibiting perception, learning, memory, model or symbol formation and manipulation, problem solving, decision making, generalization, and action in a goal-directed manner [4, 12, 16, 25, 27]. These different modes of behavior can thus be viewed as different aspects of the same process, rather than as processes of different and, possibly, isolated machines.

Chapter 30

LEARNING AUTOMATA AND ARTIFICIAL INTELLIGENCE

L. O. Gilstrap, R. J. Lee,* and M. J. Pedelty**

As noted in chapter 29, the three general approaches that have been taken for realizing the characteristics of intelligence in machines are: networks of artificial neurons, heuristic computer programs, and devices that illustrate specific learning principles, such as Walter's tortoise and Conditioned Reflex Analog (CORA) [57] and Ashby's homeostat [1].

In the network approach, selected properties of neurons are usually combined to form an artificial neuron. Van Bergeijk [7] has suggested the term neuromime to denote such analogs of nerve cells. Sets of these neuromimes are connected into networks that are then used in pattern recognition, learning, problem solving, decision making, or control applications. The neuromime network, because of its parallel action, has the advantage of high speed of operation, but the gross properties of the neuromime network depend upon the characteristics of the particular neuromimes as well as upon the patterns of interconnection of the elements.

Most of the heuristic programs for computers have emphasized problem solving or pattern recognition, rather than learning per se, although these programs generally have a learning capability. It is usually accepted that the work of Newell, Shaw, and Simon [36], reported in 1957, outlined for the first time the technique of the heuristic-programming approach to intelligent machines, although Reitman [45] has reported earlier studies of pattern recognition using computers. Because of the ready availability of computers, this approach to intelligent-machine behavior has grown rapidly, although, because of the serial operation of computers, learning or problem-solving times can be excessive for certain real-time applications.

The devices that illustrate specific learning principles are more suitable for real-time applications, but they lack the flexibility of the heuristic-programming approach. Also, most of these devices lack the generality of artificial nerve nets, and although they may serve for a

* Adaptronics, Inc., Annandale, Va.

single type of problem, such as conditioned response or pattern recognition, they usually are not suitable for many processes of machine intelligence.

In the sections that follow, more detailed descriptions of these three approaches are given, and a brief survey of the developments in these areas is provided.

Functional Properties of Neurons

The manner in which neurons handle data, search function space, store information, and become modified as a part of the learning process can provide background information for the understanding and design of neuromime models. The description of the neuron in this section is oriented to present these functional properties.

The neuron has numerous input (afferent) channels or dendrites, a cell body, a branching output (efferent) channel or axon, and synapses, where an axon branch of one neuron joins a dendrite or the cell body of the next neuron. Usually, excitation travels from a dendrite through the cell body and then along the axon, the axon usually delivering the same signal pattern to all its branches. This output signal, which may be termed explosive, is called an action potential, or spike. It is produced in the neuron and travels along the axon as a wave of activity rather than as an electric current. A stimulus applied to a neuron must be presented at or above some critical intensity, or the spike will not appear; that is, the neuron has a threshold.

A large proportion of the neurons within the brain are short compared with the sensory and motor (control) nerves that, respectively, carry information toward or away from the brain. Even within the brain there are many different types of neurons and different types of connections between them. In addition to the synapse connection, there is a functional connection called an ephapse, which exists between two neurons that are side by side. By alteration of the chemical environment of a nerve, the ephapses can be made to behave as synapses [32].

In addition to containing the neuron, the brain also contains glial, or neuroglial, cells. These constitute the central nervous system's supporting tissue, which consists mainly of cells with long fibrous branches. Though nerve cells do not multiply in the human adult and are not replaced when destroyed, certain neuroglial cells have the capability for active proliferation. Some recent observations that neuroglial cells grown in tissue culture produce slow action potentials suggest caution in assuming that these cells are restricted to mechanical roles [8].

In the neuron the magnitude of the spike (if it occurs at all) is essentially independent of the size of the stimulus; any stimulus above

the threshold generally produces the full spike magnitude. The fact that the spike appears not to vary with the size of the stimulus has led to the term all or none in reference to the single spike. When a neuron so responds, it is said to fire. The spike is triggered by the stimulus, but the energy for the spike and for its transmission along the axon comes from the neuron itself; so the neuron may be considered as an active rather than a passive electrical element. The magnitude and propagation speed of the spike do vary with such factors as the diameter of the fiber, the past history of the neuron, the oxygen supply, drugs that may be present, and temperature.

If two stimuli in succession are presented to the neuron and if the first stimulus was adequate for firing the neuron, the threshold to the second stimulus depends on how closely it follows the first. If the second stimulus follows too closely on the first, it will produce no effect. Then, after a slightly longer period, the neuron can be fired by the second stimulus, but the result will be a smaller spike, a slower propagation speed, and a higher threshold than for the first stimulus [10]. These two periods are called the absolute refractory period and the relative refractory period, respectively. After the relative refractory period, a longer period occurs in which the neuron has a lower threshold than it had to the first stimulus. Then a still longer period occurs in which the neuron has a very slightly higher threshold than it had to the first stimulus. This latter period is enhanced by prolonged stimulation and fatigue [32].

If two stimuli in succession are presented to the neuron during a short period, called the period of latent addition, and if neither stimulus separately would have been adequate to fire the neuron, then the two stimuli may fire the neuron. This may be termed the summation of input signals. It may occur either for signals at a single synapse or for signals coming in at different synapses or different dendrites [32]. Often a signal coming into a synapse will have the effect of preventing rather than augmenting the firing of the neuron; this is inhibition. As a result of the effect of summation of input signals and inhibition, the neuron can function as a logic gate. Also, delay will occur at the synapse [32] and can occur as a continuation of firing after cessation of the stimulus [11].

The neuron response has been referred to as "all or none," which implies digital operation. However, it has been seen that the past history of the neuron and other factors can influence the amplitude of the spike, the propagation speed of the spike along the axon, and the length of the refractory period. These are continuous, or analog, rather than discrete, or digital, quantities. A very important analog parameter of operation of the neuron is the spike repetition rate, which can be

influenced by the intensity of the stimulus in the case of sensory neurons. This is the form of analog signal coding that was called a pulse-density system by von Neumann [34]. A related form of analog signal coding is in terms of the number of neurons that will be fired by a given signal. If various neurons receiving a stimulus have different thresholds, a larger stimulus will affect more neurons.

Adaptation is an analog effect that occurs in a neuron when there is a change in an input signal; it may be viewed as a form of decay involving a time constant. Thus, consider a neuron that is stimulated by the appearance of a continuing signal, such as would be produced for a light-sensitive neuron presented with a continuing light stimulus. The result may be that the neuron will erupt into a burst of activity that then gradually dies out until the neuron becomes dormant again. The dying out of the bursts is in terms of a decreasing spike repetition frequency. This function is roughly comparable with that of a resistance-capacitance phase-lead network (high-pass filter). In some neurons, a burst of activity that then dies out will occur only when the light goes off. One may also find various other neuron functions described in the literature [11, 32].

The neuron can thus be thought of as operating in both digital and analog modes, and certain parameters of its functioning may be affected by various conditions other than its instantaneous stimulus.

Some doubt exists as to whether neurons in the adult can form new synapses with other neurons as a means of learning and adaptation. Seemingly this type of change, internal changes in the neuron, changes in ephapses, or changes in the neuroglial cells must occur to account for storage of information in the animal learning process, unless, as is unlikely, all information storage is in terms of impulses in circulating loops of neurons. For example, Norbert Wiener has pointed out that:

> Some definite combinations of impulses on the incoming neurons having synaptic connections with a given neuron will cause it to fire, while others will not cause it to fire. This is not to say that there may not be other, non-neuronic influences, perhaps of a humoral nature, which produce slow, secular changes tending to vary that pattern of incoming impulses which is adequate for firing [59, p. 142].

He goes on to introduce the term "affective tone" and to suggest:

> The essential thing is this: that affective tone is arranged on some sort of scale from negative « pain » to positive « pleasure »; that for a considerable time, or permanently, an increase in affective tone favors all processes in the nervous system that are under way at the time, and gives them a secondary power to increase affective tone; and that a decrease in affective tone tends to inhibit all processes under way at the time, and gives them a secondary ability to decrease affective tone [59, p. 150].

Actually, electrical as well as humoral signals to neurons could produce internal changes in the neuron to favor existing processes. According to D. O. Hebb, assemblies of neurons may become organized through their interactions, based upon changes that occur within the neurons because of electrical stimulations from other neurons. He postulated that:

> *When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased* [18, p. 62].

The effect of such a process in promoting organization in an originally unorganized group of neurons was simulated on a computer by Rochester et al. [46], with some positive and some negative results.

An essential factor in general learning is a directive principle, such as an overall goal, which has already been mentioned. This would promote organization of groups of neurons. Wiener's "pleasure," "pain," and "tone," mentioned above, satisfy this requirement; the lack of this type of mechanism in Hebb's work may account for the incomplete success of his theories.

Another factor in learning is the degree to which statistical operation or spontaneity enters into the operation of a neuron. Spontaneity is an important property of human behavior. It could arise from thermal noise or chemical fluctuations in the neurons. A learning device can be characterized by searching function space and then by retention of the most satisfactory functions found. Searching of the function space in the human brain could be based on spontaneous neural activity. Spontaneous neural discharges are discussed by Bullock [11] and Jung [22].

Specific Neuromimes and Other Building Blocks

In 1943 Warren McCulloch and Walter Pitts showed that the propositional calculus (e.g., Boolean algebra) was sufficient to describe the gross behavior of any nerve net [30]. That is, given such a net with observed behavior, there is another net whose behavior is the same (perhaps in a different time scale) and whose individual events (the presence of a pulse at a given time and place) and their relations are describable in terms of the propositional calculus. Thus was born the Pitts-McCulloch formal neuron.

Other mathematical tools pertinent to the design of intelligent machines are conditional probability [43, 53 to 56] and inductive inference. The changes of inputs and outputs of an automaton are spoken of as events. The presence or absence of a binary event may be

denoted by a literal or its complement (for example, A, A'). The same notation may be used for classes of events, and the classes may be visualized in terms of generally overlapping point sets. Therefore, an event (set of input or output states) is, or is not, a member of Class A, for example. This same event may also be a member of Class B, depending on the definition of that class. If every member of A were some member of B, A implies B and $p\ (A|B)$, the probability of a member of A, given a member of B, could be computed.[1] If the latter were 1, then B implies A and the classes are equivalent. Implications may also exist among temporally disjointed events; e.g., if A is preceded or followed by B, again A implies B. If A is not always accompanied by B, A probably implies B and $p\ (B|A)$ may be computed. The same notions apply to the selection of a response (a certain input event implies a certain appropriate output event) and to the formation of abstractions (e.g., all positive first-order differentials imply an increase in pressure). Thus, a sophisticated problem-solving mechanism must be capable of computing the truth values of (probabilistic) implications or, equivalently, conditional probabilities. Most of the building blocks to be described are capable of forming nets exhibiting these properties.

Pitts-McCulloch neurons and some other devices called plastic neurons, majority organs [39], and Perceptron association units are all threshold devices that depend on linear (weighted) summation of the input signals; i.e., the neuromime fires if and only if $\Sigma_i a_i x_i \geqq T$, where the a_i and T are arbitrary real numbers representing the signal strength of the ith connection and the threshold, respectively, and x_i is a binary variable representing the instantaneous state of the ith input lead. The a_i's are made greater than or less than 0, corresponding to excitation and inhibition, respectively. Obviously, such a device may produce the functions of AND, OR, and NEGATION, and thus the algebra of their nets is functionally complete.

With the exception of a type of Pitts-McCulloch neuron, to be described, these devices all realize the set of linearly separable Boolean functions having input weights that are solutions of the inequality given in the preceding paragraph. They incorporate no delay or refractory mechanism per se, and the design of their nets [23] consists mainly of forming switching circuits depending on a partitioning of the mappings into linearly separable functions [33, 61].

The major advantage of the plastic neurons [60] and the Perceptron [48] association units lies in their ability to form variable switching functions by having the weights a_i depend on the previous history of the net. The weights are often also under the control of the experimenter. In this way, for example, a net of plastic neurons can be made to form

[1] The quantity $p(B|A)$ is unity.

classes of input patterns without the designer's having to specify a priori the types of patterns to be encountered [40].

Recently a Pitts-McCulloch model has been proposed that may have interacting inhibition among inputs, i.e., a pulse on one input may block a pulse on some other specific input or inputs. This device is therefore universal inasmuch as it can compute any of the 2^{2^n} switching functions of n variables [9].

On the basis of Hebb's postulate concerning the facilitation of neural pathways, several workers [15, 20, 46] have made nerve-net models that are unlike the logical devices previously discussed. The neuromimes, or cells, postulated in these theories possess a refractory period and may possess a delay. Their nets have many feedback connections, so that reverberating chains may be set up. The nets are arranged so that the activity of one set of cells may inhibit activity in another set. These sets form spontaneous assemblies oscillating at some natural frequency. Some of the studies have incorporated a type of reinforcement learning.

A most important set of special considerations in the study of neuromime networks is the relation of such networks to their environment. If the automaton is to interact in a natural rather than an artificial environment, it must include sensors, a supply of energy, and means for acting upon the environment. Inasmuch as the environment of the automaton may include a man, this point is central to the problem of the interaction between automaton and man and involves such questions as the following:

1. Can the automaton learn to evaluate its own input information, or will all input information have to be evaluated or interpreted for it by a human operator?

2. Can an automaton follow general rather than specific instructions given by the operator?

3. Can the automaton proceed to develop its own methods for solving a problem that it has encountered without having to call on the operator for help?

4. Can the automaton teach itself in carrying out the assignments it is given and thus develop its capability through its own interaction with an inanimate environment, or will it require a human trainer in order to learn?

5. Can the automaton learn to provide optimum interaction with a particular operator by, in effect, adapting to the characteristics of that operator?

Certainly, complete answers to such questions would depend upon the type of problem or task involved. For simple problems the answers to these questions are clearly affirmative, and the trends of current research indicate that the capability of learning automata will be extended

toward more complex problems and toward more complete generality of function.

In 1952 a building-block device, called a Reron, was developed by Lee [16, 27] for use as a generalized learning element for automata operating in either natural or artificial environments. Learning in a Reron consisted of trial-and-error selection of a Boolean function of two inputs to the Reron. To demonstrate that a network of Rerons could function in a natural environment, Lee hypothesized a tortoise-like vehicle that employed Rerons as its decision elements. Motivated by such goals as maintaining a charge on an internal battery and maintaining temperature within tolerable limits, this vehicle was shown to be capable of learning to navigate through T mazes, to pursue similar vehicles, to produce and respond to meaningful sounds, and to perform other tasks. Lee's subsequent Artron [13, 24, 26, 29] was evolved from the Reron but has a wider repertory of Boolean functions. A more advanced device, the Neurotron, to be described in brief detail below, has been developed more recently by Lee and features a combination of analog and Boolean functions that are subject to learning.

The Neurotron may be thought of as having two input channels and one output channel and, through learning, may develop any of the 16 possible logical functions of the input variables. It may also learn to produce a phase lead or lag, to control gain, to produce signal inversion, and to perform, to a limited extent, waveform shaping. In a network of several Neurotrons, adaptive pulse delay, independent adaptive delay of the leading or trailing edge of a pulse, adaptive peak detecting, adaptive generation of a pulse train or sine wave, and various other functions can be obtained.

In the Neurotron, the adaptive analog processes are important, especially in the formation of functions for control systems, since the direct use of analog functions is much simpler than the simulation of analog processes by digital means. The adaptive digital and logical processes, on the other hand, are particularly important in problems involving decision making.

Although either a random or a systematic network containing a sufficient number of Neurotrons will generate any function of its inputs, it is noted that, if the network is expected to operate in a manner similar to that of the central nervous system, the number of input channels with signals will normally be much less than the total number of input channels to the network. This will require far fewer Neurotrons than if all functions of the inputs to the network were required, even though it is not known a priori which subset of input channels will carry signals.

The Neurotron search mode in function space is based upon random signals that are controlled in average amplitude. These signals come

independently to each Neurotron and provide spontaneity in searching function space. The random signals are, respectively, decreased or increased in average amplitude by control signals called reward or punishment signals. Reward increases the likelihood that existing or recent functions will persist; punishment reduces this likelihood. The reward-punishment mechanism serves to select and retain useful trial functions that have occurred. This selection and retention constitute the learning process. Although reward and punishment come in parallel to all Neurotrons, the effect can be different in each, for the effect of reward and punishment depends on the past history of the Neurotron and on the existing internal conditions. The history and existing conditions depend upon the detailed search mode and on the input signals, which have been different in the different Neurotrons. The reward and punishment are provided by generalized goal circuits that are prespecified by the designer and that can be further built up by learning. The goal circuits built up by learning may be specialized to the problem at hand and can thereby serve to increase the learning speed.

Learning in the Neurotron or Neurotron network normally occurs by small steps, probably the only practical way in which complex behavior can arise through learning. The statistical character of the search mode, however, leaves open the possibility of taking occasional large steps. The occurrence of a trial step in function space does not itself alter the conditioning of the Neurotron system; it alters only the function of the system. Conditioning is stored in terms of the future probability or likelihood of the Neurotron's being in a particular state, i.e., of having a particular function, in addition to the present state or function itself being stored. In this sense, the Neurotron can be classed as a probability-state neuromime.

Reward and punishment affect the future probabilities of a particular Neurotron function; these probabilities, in turn, govern subsequent search activity and thus subsequent changes in functions. As a limit of the learning process, all such probabilities can approach unity, so that search activity will cease; however, if environmental conditions change, the system can always develop new probabilities and new functions within the Neurotrons.

John Holland [19] has suggested a universal computer of modular construction, which, although not a neuromime, is nevertheless governed, in its network organization, by rules analogous to those found applicable to artificial nerve nets. Its major advantages over conventional computers are (1) parallel processing (i.e., the simultaneous execution of a large number of subprograms) and (2) self-organization. This concept combines many of the advantages of heuristic programs, nerve nets, and special devices such as the homeostat. A single module has operation

codes analogous to those in a conventional computer. It also has codes that cause opening of processing channels via its neighbors. These channels may specify the data address or the module that is to be the successor. Operation starts at a set of modules and proceeds to each of their successors. Since the contents of module registers can be changed, the program has the potentiality of modifying itself radically. By changing channels, modules can be added to or subtracted from a subprogram. The success of a subprogram leads to reinforcement. Blind variation can be introduced by injecting noise into the registers.

Crane has developed a device composed of thermistor material and a distributed capacitance, called the Neuristor, which can serve as a building-block component and which has functions much like those of a neuron [14]. It has a threshold and a refractory period, and it propagates a spike or action potential, the energy for the spike and for its transmission coming from the Neuristor. This neuromime does not itself learn, but it can be used to build up systems that learn.

Another device in which signal propagation is maintained by distributed energy is Stewart's MABIAC. In this device excitation is propagated and processed in three dimensions in a system of iron balls in acid [51]. One may draw a rough comparison between the system and an entire brain, with the lattice points in the configuration serving as building-block elements. The system can learn by the punishment of recent erroneous responses and thus foster new behavior. Stewart has proposed a mechanism whereby only recently active functional elements are affected by a gross indication of unsatisfactory behavior [51].

Devices Illustrating Specific Learning Principles

Probably the earliest learning machine to be described was a hydraulic device described by Russell in 1913 [49]. The device employed a learning process similar to a combination of conditioned reflex and habit used to set up switching functions. Habit learning in Russell's machine is illustrated by the facilitation of the first response to a given stimulus in future applications of the given stimulus, and conditioned reflex is illustrated by evocation of a particular learned response to a pattern by isolated elements of that pattern.

Walter has described mechanisms that simulate a number of learning principles [57]. The Conditioned Reflex Analog (CORA), when incorporated into *Machina speculatrix*, a mechanical tortoise with several unconditioned reflexes, resulted in a device clearly exhibiting an analog of conditioned-reflex behavior. When CORA was connected to an obstacle-avoiding mechanism in *M. speculatrix*, the machine could be taught that sound means trouble. The training process consisted in blowing a whistle

and kicking the shell a few times. After this the device "backed away and turned from an 'imagined' obstacle" at each whistle sound.

The principle of ultrastability was demonstrated by Ashby [1 to 3] in a device termed a homeostat. This device modifies its internal parameters in such a way as to maintain its outputs within fixed limits independent of various contingencies in the inputs to the homeostat. The principle of ultrastability and that of multistability, which was derived from it, have been shown to be important in the adaptation process.

Barlow [5] has discussed the role of redundancy reduction in perception and has described a redundancy-reduction mechanism that learns to code two binary inputs into two outputs so that the output redundancy is less than the input redundancy. The inputs are correlated in terms of their past history, and in a consistent environment the redundancy is removed from the outputs.

A growing system capable of developing sensory receptors "for those variables which are sensed with advantage" has been described by Pask [38]. This elegant system also models aspects of the process of evolving successively more complex mechanisms in a nonteleological manner and with aspects of individuality. It represents an attack on some of the more fundamental questions of self-organizing systems.

The adaptive control systems [17, 52] that have been widely developed and applied, primarily to flight-control problems, have, in general, a limited learning capability. These systems were originally restricted to performance modifications through gain and time-constant changes. More recently there have been attempts made to bridge the gap between adaptive-control theory and communications theory by Bellman [6] and the gap between adaptive-control theory and neural-network models by Pedelty [39].

Heuristic Computer Programs

Typically the procedure for preparing a heuristic problem-solving program is as follows: The programmer observes the activities of a human being solving a problem, which may be a problem in musical composition, proving a theorem in logic, trigonometry, geometry, or algebra, or a game such as chess [35]. The programmer records each step taken by the problem solver and such comments or explanations as may be provided. The record of the problem solver's activity is analyzed to determine activities or processes and the rules of employing these processes used by the human problem solver. This determination of processes is inductive. When the programmer has decided that he has defined a set of processes and heuristics that will account for the observed activity, a computer program is written to perform similar activities. Reitman [44,

45] has given descriptions of the technique of heuristic programming for the problem of composing music, along with surveys of the heuristic-programming field.

The above description is highly oversimplified in that the programmer may have to observe the problem solver many times before he can formulate adequate hypotheses about the processes involved. Likewise the programmer may be only partially successful with his first program, and many parts of the program may have to be rewritten.

A more general heuristic program has been developed by Newell, Shaw, and Simon [37]. This program, called the General Problem Solver (GPS), has the capability of learning how to solve problems. It consists of two distinct programs: the performance program, which solves problems, and the learning program, which modifies the performance program to obtain better results in problem solving. The learning program was developed in a manner similar to that for other heuristic programs, and it is general in that it can be used with a variety of performance programs. The General Problem Solver can, in principle, be used with any problem that can be defined in terms of objects, operators for altering objects, properties of objects, and differences between objects. In the General Problem Solver, as in other heuristic programs, the representations of raw sensory data, i.e., the objects, are not learned by the program and must be supplied. In this sense, the heuristic programs thus far do not "start from scratch" in the learning process.

Samuel has reported a checker-playing computer program that employs a redundant and incomplete list of parameters that are thought to be relevant to the game of checkers [50]. These parameters, derived from the advice of experts in checker playing, are evaluated for relative merit by the computer program in the course of actual play and are continually readjusted. In a relatively short time this program was able to learn to play well enough to outplay the person who wrote the program.

COMPARISONS OF AUTOMATA

The neuromime network, the heuristic-programming system, and the devices that illustrate specific learning principles may be compared in various ways. Learning may be generalized as in the Neurotron system or the General Problem Solver, or it may be specialized as in adaptive servos. In the former case, emphasis is on performance; in the latter case, emphasis is on optimization, although the more general neuromime networks can serve as adaptive servos. All these approaches are currently the subject of considerable study.

It is interesting to compare the probability-state neuromime network

with the heuristic computer. Because of the parallel operation of the neuromime network and the sequenced operation of the heuristic-computer program, the neuromime system is theoretically much faster than the computer approach. On the other hand, the computer, because it time-shares equipment, may realize some theoretical size advantage. The neuromime system enjoys an inherent reliability advantage over both the computer approach and the specialized approaches considered earlier because it is inherently self-correcting and parallel in operation. The computer approach, on the other hand, has far more flexibility for research purposes, and, in fact, neuromime systems are normally simulated on a computer for research purposes.

An interesting difference that exists between the more advanced neuromime systems, such as the Neurotron system, and the more advanced heuristic computer systems, such as the General Problem Solver, is that the neuromime systems begin with raw input or sensory data, develop their own definitions and interpretations of the environment, and solve problems in these terms, whereas the heuristic-computer program must have the objects in the environment defined for it. The performance program of the heuristic-computer system corresponds to the neuromime network, and the learning program of the heuristic-computer system corresponds to the goal system used with the network.

The learning process will now be illustrated with a specific example of automata. Two types of automata will be considered, an automaton A that searches for new functions according to a fixed search pattern, and an automaton B that searches in the same environment by using a random search. The second automaton may be said to have spontaneity, and it will be seen to be more satisfactory than the first. An abstract environment is defined as in Figure 30-1,[2] in which 1:1, 0:1, 1:0, or 0:0 denote an input-output combination $(X{:}Y)$ for the environment. That is, the input to the environment from the machine is the first digit, and the output from the environment to the machine is the second digit. A change in state of the environment is indicated by a directed line, the states of the environment being a, b, c, or d. This environment has memory, its outputs being functions not only of its inputs but also of its existing state. Furthermore, the next state of the environment will be a function of inputs to the environment and its

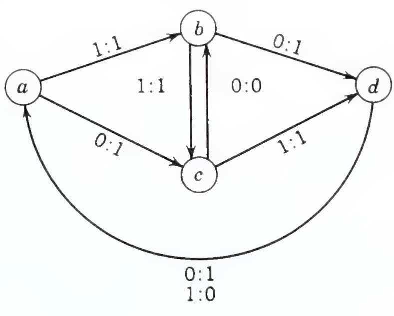


FIG. 30-1 *Transition diagram of a sequential environment.*

[2] This type of diagram is discussed in general by Phister [41].

existing state. It is assumed that time is discrete, with no change of the environmental state occurring between time steps. The output of the automaton, including the all-zero output, can cause a change in the environment.

As an example, let the goal of the automaton be to produce an output Y from the environment, which is an infinite series of 1's. This series may be the output of a success-measuring device, the symbol 1 representing a successful outcome. Let the experiment consist in having the automaton produce equal-length sequences of output symbols X, which are the inputs to the environment. Let automaton A be defined as one utilizing an error-correction scheme that inverts those bits in X which correspond in sequential position to zeros in Y during any given trial. If the automaton has a three- or four-length sequence, certain combinations of initial automaton output sequences will lead rapidly to a correct solution.

For example, if the initial state of the environment is a and if the initial automaton output sequence is 000, then the first trial would be: 0:1 (a to c); 0:0 (c to b); 0:1 (b to d). The automaton stops in environment state d and detects the error in the second bit. By means of the error-correction scheme, which inverts those bits in X which correspond to zeros in Y, the next sequence will be 010. The second trial is therefore 0:1 (d to a); 1:1 (a to b); 0:1 (b to d). As can be seen, automaton A can now produce the infinite series of 1's by infinitely repeating the sequence 010.

However, if the automaton had started in environment state a and with output sequence 001, the results for the first trial would be: 0:1 (a to c); 0:0 (c to b); 1:1 (b to c). The error in the second bit would be changed, and the new sequence would be 011. The second trial would be: 0:0 (c to b); 1:1 (b to c); 1:1 (c to d). The error now would be in the first bit, and therefore the third trial would use sequence 111 and would result in: 1:0 (d to a); 1:1 (a to b); 1:1 (b to c). Once again the error would be in the first bit and would result in the new sequence 011 for the fourth trial: 0:0 (c to b); 1:1 (b to c); 1:1 (c to d). Clearly, unless automaton A were able to change the routine, it would now continue to oscillate indefinitely from one of the last two trials to the other.

In automaton B all bits in its output sequence will be liable to change (according to the Monte Carlo statistics) if any environmental output is a 0, but the automaton will keep the same sequence if all three environmental outputs are 1's. This scheme may seem more haphazard than the scheme for automaton A, and, indeed, it will generally lead to a longer learning time if the initial environment state and automaton output sequence are a and 000, respectively. On the other hand, it will always lead to a solution in the environment of Figure 30-1, given enough trials; it will always lead to a breakout of the type of uselessly repetitive cycling illustrated above for automaton A for initial environment state and automaton

output sequence *a* and 001, respectively. Actually, the number of trials required for automaton *B* in the environment of Figure 30-1 is not great. Two Monte Carlo simulations for initial environment states and automaton output sequences *a* and 001, respectively, are shown in Table 30-1. A tossed coin was used as the "Monte Carlo generator" for Table 30-1 and provided

TABLE 30-1 *Two Monte Carlo Simulations for Initial Environment State and Automaton Output Sequence a and 001, Respectively*

Automaton output sequence	Trial	Environment stops in state
1st run:		
001	0:1 (*a* to *c*); 0:0 (*a* to *b*); 1:1 (*b* to *c*)	*c*
010	0:0 (*c* to *b*); 1:1 (*b* to *c*); 0:0 (*c* to *b*)	*b*
111	1:1 (*b* to *c*); 1:1 (*c* to *d*); 1:0 (*d* to *a*)	*a*
111	1:1 (*a* to *b*); 1:1 (*b* to *c*); 1:1 (*c* to *d*)	*d* (reward but no solution)
111	1:0 (*d* to *a*); 1:1 (*a* to *b*); 1:1 (*b* to *c*)	*c*
010	0:0 (*c* to *b*); 1:1 (*b* to *c*); 0:0 (*c* to *b*)	*b*
111	1:1 (*b* to *c*); 1:1 (*c* to *d*); 1:0 (*d* to *a*)	*a*
000	0:1 (*a* to *c*); 0:0 (*c* to *b*); 0:1 (*b* to *d*)	*d*
000	0:1 (*d* to *a*); 0:1 (*a* to *c*); 0:0 (*c* to *b*)	*b*
011	0:1 (*b* to *d*); 1:0 (*d* to *a*); 1:1 (*a* to *b*)	*b*
101	1:1 (*b* to *c*); 0:0 (*c* to *b*); 1:1 (*b* to *c*)	*c*
100	1:1 (*c* to *d*); 0:1 (*d* to *a*); 0:1 (*a* to *c*)	*c* (solution)
2d run:		
001	0:1 (*a* to *c*); 0:0 (*c* to *b*); 1:1 (*b* to *c*)	*c*
001	0:0 (*c* to *b*); 0:1 (*b* to *d*); 1:0 (*d* to *a*)	*a*
001	0:1 (*a* to *c*); 0:0 (*c* to *b*); 1:1 (*b* to *c*)	*c*
011	0:0 (*c* to *b*); 1:1 (*b* to *c*); 1:1 (*c* to *d*)	*d*
111	1:0 (*d* to *a*); 1:1 (*a* to *b*); 1:1 (*b* to *c*)	*c*
010	0:0 (*c* to *b*); 1:1 (*b* to *c*); 0:0 (*c* to *b*)	*b*
000	0:1 (*b* to *d*); 0:1 (*d* to *a*); 0:1 (*a* to *c*)	*c* (reward but no solution)
000	0:0 (*c* to *b*); 0:1 (*b* to *d*); 0:1 (*d* to *a*)	*a*
010	0:1 (*a* to *c*); 1:1 (*c* to *d*); 0:1 (*d* to *a*)	*a* (solution)

blind variation. Selective survival was provided by keeping the same output sequence if all three environmental outputs were 1's; that is, the output sequence that yields 1's from the environment was the sequence that survived.

A single Neurotron, because of its ability to make random changes in its goal-seeking tactics, is capable of providing a transfer function comparable with that of automaton *B*. One Neurotron and a goal circuit that punishes the Neurotron when a zero occurs but rewards it when no zero occurs in the sequence of 3 bits from the environment could serve as the entire automaton *B*, depending on the reward and punishment parameters

of the device. A network of Neurotrons would be capable of learning to succeed in various environments with characteristics comparable with but differing from those shown in Figure 30-1. The network would also be able to select the length of its sequence so that it would not be limited to a sequence of 3 bits. Thus, its searching of function space would include decisions as to sequence length and content.

SUMMARY

Some underlying principles that unify the field of machine intelligence have been discussed, but certainly it cannot be said that all such principles are completely delineated as yet. To this extent this field presents a changing complexion. However, there no longer appears to be any question of whether or not a machine can perform tasks such as pattern recognition, learning, and problem solving. The practical questions now appear to be directed at how to perform such tasks to some reasonable degree of complexity or within a reasonable time. It also appears safe to say, at this time, that theoretical work will continue to be concerned with goals and goal-directed behavior as one of the central problems in the field, as well as with the strategies for selecting new forms of behavior. The problems associated with goals and search patterns in machines are related to the broader philosophical problem of human values, and it is to be hoped that developments in the machine-intelligence field may stimulate advances in the broader problem.

BIBLIOGRAPHY

1. Ashby, W. R.: "Design for a Brain," John Wiley & Sons, Inc., New York, 1952.

2. ———: Homeostasis, *Cybernetics, Trans. Ninth Conf. Josiah Macy, Jr., Foundation,* pp. 73–108, March, 1952.

3. ———: "An Introduction to Cybernetics," Chapman & Hall, Ltd., London, 1957.

4. ———: What is an Intelligent Machine? *Proc. WJCC,* pp. 275–280, 1961.

5. Barlow, H. B.: Sensory Mechanisms, the Reduction of Redundancy, and Intelligence, *Proc. Symposium Mechanization Thought Processes,* vol. 2, pp. 535–559, 1958.

6. Bellman, R.: "Adaptive Control Processes: A Guided Tour," Princeton University Press, Princeton, N.J., 1961.

7. Bergeijk, W. A. van: Nomenclature of Devices Which Simulate Biological Behavior, *Science,* vol. 132, pp. 1248–1249, Oct. 28, 1960.

8. Best, C. H., and N. B. Taylor: "The Physiological Basis of Medical Practice," 7th ed., The Williams & Wilkins Company, Baltimore, 1961.

9. Blum, M.: Properties of a Neuron with Many Inputs, in Bionics Symposium, *USAF WADD Tech. Rept.* 60-600, pp. 55–82, September, 1960. (NOTE: Section 1.7 should read "If the neuron makes *no* error . . . .")

10. Brazier, M. A. B.: "The Electrical Activity of the Nervous System," Pitman Publishing Corporation, New York, revised 1953, reprinted 1958.

11. Bullock, T. H.: Neuron Doctrine and Electrophysiology, *Science*, vol. 129, pp. 997–1002, Apr. 17, 1959.

12. Campbell, D. T.: Blind Variation and Selective Survival as a General Strategy in Knowledge-processes, in M. C. Yovits and S. Cameron (eds.), "Proceedings of the Interdisciplinary Conference on Self-organizing Systems," pp. 205–231, Pergamon Press, Inc., New York, 1960.

13. Corneretto, A.: Electronics Learns from Biology, *Electronic Design,* vol. 8, pp. 43–45, Sept. 14, 1960.

14. Crane, H. D.: The Neuristor, *IRE Trans. on Electronic Computers,* vol. EC-9, pp. 370–371, September, 1960.

15. Farley, B. G., and W. A. Clark: Simulation of Self-organizing Systems by Digital Computers, *IRE Trans. on Inform. Theory,* vol. PGIT-4, pp. 76–84, 1954.

16. Gilstrap, L. O., and R. J. Lee: Learning Machines, in Bionics Symposium, *USAF WADD Tech. Rept.* 60-600, pp. 437–450, September, 1960.

17. Gregory, P. C. (ed.): Proceedings of the Self Adaptive Flight Control Systems Symposium, *USAF WADC Tech. Rept.* 59-49, ASTIA AD-209 389, March, 1959.

18. Hebb, D. O.: "The Organization of Behavior," John Wiley & Sons, Inc., New York, 1949.

19. Holland, J. H.: Iterative Circuit Computers, *Proc. WJCC,* pp. 259–266, 1960.

20. ——— and W. Crichton: A New Method of Simulating the Central Nervous System Using an Automatic Digital Computer, *Univ. Mich. Rept.* 2144-1195-M, March, 1959.

21. Householder, A. S., and H. D. Landahl: "Mathematical Biophysics of the Central Nervous System," Principia Press, Inc., Bloomington, Ind., 1945.

22. Jung, R.: Neuronal Integration in the Visual Cortex and Its Significance for Visual Information, in Walter A. Rosenblith (ed.), "Sensory Communication," pp. 627–674, Massachusetts Institute of Technology Press, Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1961.

23. Keller, H. B.: Finite Automata, Pattern Recognition and Perceptrons, *J. Assoc. Computing Mach.,* vol. 8, pp. 1–20, 1961.

24. Klass, P. J.: Self-learning Machines to Use Artrons, *Aviation Week,* vol. 74, pp. 52–53, Feb. 20, 1961.

25. Kleene, S. C.: Representation of Events in Nerve Nets and Finite Automata, in C. E. Shannon and J. McCarthy (eds.), "Automata Studies," pp. 3–41, Princeton University Press, Princeton, N. J., 1956.

26. Lee, R. J.: Generalization of Learning in a Machine, *Preprints Papers Presented at Fourteenth Natl. Meeting ACM,* pp. 21-1–21-4, Sept. 1-3, 1959.

27. ———: A Generalized Learning-machine, privately published, copyright Dec. 20, 1954. (Copies may be obtained from the author.)

28. ———: Internal Circuitry of a Reron, privately published, copyright June 13, 1955, Library of Congress Card TK 7882.C5L4. (Copies may be obtained from the author.)

29. ———: More on Bionics and Artron, *Electronic Design*, vol. 9, pp. 166–167, June 21, 1961.

30. McCulloch, W. S., and W. Pitts: A Logical Calculus of the Ideas Immanent in Nervous Activity, *Bull. Math. Biophys.*, vol. 5, pp. 115–133, 1943.

31. MacKay, D. M.: The Epistemological Problem for Automata, in C. E. Shannon and J. McCarthy (eds.), "Automata Studies," pp. 235–251, Princeton University Press, Princeton, N.J., 1956.

32. Morgan, C. T., and E. Stellar: "Physiological Psychology," chap. 4, McGraw-Hill Book Company, Inc., New York, 1950.

33. Muroga, S., I. Toda, and S. Takasu: Theory of Majority Decision Elements, *J. Franklin Inst.*, vol. 271, pp. 376–418, May, 1961.

34. Neumann, J. von: "The Computer and the Brain," Yale University Press, New Haven, Conn., 1958.

35. Newell, A., J. C. Shaw, and H. A. Simon: Chess-playing Programs and the Problem of Complexity, *IBM J. Research Developm.*, vol. 2, pp. 320–335, October, 1958.

36. ———, ———, and ———: Empirical Explorations with the Logic Theory Machine: A Case Study in Heuristic, *Proc. WJCC*, pp. 218–230, February, 1957.

37. ———, ———, and ———: A Variety of Intelligent Learning in a General Problem Solver, in M. C. Yovits and S. Cameron (eds.), "Proceedings of the Interdisciplinary Conference on Self-organizing Systems," pp. 153–189, Pergamon Press, Inc., New York, 1960.

38. Pask, G.: The Natural History of Networks, in M. C. Yovits and S. Cameron (eds.), "Proceedings of the Interdisciplinary Conference on Self-Organizing Systems," pp. 233–263, Pergamon Press, Inc., New York, 1960.

39. Pedelty, M. J.: Introducing the Neuron, *Control Eng.*, vol. 9, pp. 91–94, April, 1962, and Neuron Technology Builds Reliable Circuits, *Control Eng.*, vol. 9, pp. 115–119, May, 1962.

40. ———: A Memory Complex for an Adaptive Control System, in F. Caruthers and H. Levenstein (eds.), "The Adaptive Control Symposium," Pergamon Press, Inc., New York, in press.

41. Phister, M.: "Logical Design of Digital Computers," John Wiley & Sons, Inc., New York, 1958.

42. Rashevsky, N.: "Mathematical Biophysics," Dover Publications, New York, 1960.

43. Reid, K. H.: Design Studies of Conditional Probability Computers, paper presented at Second Bionics Symposium, Cornell University, Ithaca, N.Y., September, 1961.

44. Reitman, W. R.: Programming Intelligent Problem Solvers, *IRE Trans. on Human Factors in Electronics*, vol. HFE-2, no. 1, pp. 26–33, March, 1961.

45. ———: Information Processing Languages and Heuristic Programs: A New Stage in the Bead Game, in Bionics Symposium, *USAF WADD Tech. Rept.* 60-600, pp. 409–417, September, 1960.

46. Rochester, N., et al.: Tests on a Cell Assembly Theory of the Action of the Brain, Using a Large Digital Computer, *IRE Trans. on Inform. Theory,* vol. PGIT-2, pp. 80–93, 1956.

47. Rosenblatt, F.: The Perceptron, *Cornell Aeronaut. Lab. Rept.* VG-1196-G-1, January, 1958.

48. ———: Perceptron Simulation Experiments, *Proc. IRE,* vol. 48, pp. 301–309, March, 1960.

49. Russell, S. B.: A Practical Device to Simulate the Working of Nervous Discharges, *J. Animal Behavior,* vol. 3, pp. 15–35, 1913.

50. Samuel, A. L.: Some Studies in Machine Learning Using the Game of Checkers, *IBM J. Research Developm.,* vol. 3, pp. 210–229, July, 1959.

51. Stewart, R. M.: Peripheral Access Lattices, Space Electronics Corp. (Now Space-General Corp.), Glendale, Calif., Oct. 11, 1960.

52. Truxal, J. G.: Congress of International Federation on Automatic Control, *Ordnance Research Bull.,* vol. ORB 60-8, pp. 1–4, August, 1960.

53. Uttley, A. M.: Conditional Probability Computing in a Nervous System, in "Mechanization of Thought Processes," vol. 1, pp. 119–147, H.M. Stationery Office, London, 1959.

54. ———: Conditional Probability Machines and Conditioned Reflexes, in C. E. Shannon and J. McCarthy (eds.), "Automata Studies," pp. 253–276, Princeton University Press, Princeton, N.J., 1956.

55. ———: Conditional Probability as a Principle of Learning, *Proc. First Intern. Congr. Cybernetics,* pp. 830–856, 1956.

56. ———: The Design of Conditional Probability Computers, *Inform. & Control,* vol. 2, pp. 1–24, April, 1959.

57. Walter, W. G.: "The Living Brain," W. W. Norton & Company, Inc., New York, 1953.

58. ———: A Machine That Learns, *Sci. American,* vol. 185, no. 2, pp. 60–63, August, 1951.

59. Wiener, N.: "Cybernetics," John Wiley & Sons, Inc., New York, 1948.

60. Willis, D. G.: Plastic Neurons as Memory Elements, *Proc. Intern. Conf. Inform. Processing, UNESCO,* pp. 290–298, 1959.

61. Winder, R. O.: Some Recent Papers in Threshold Logic, *Proc. IRE,* vol. 49, p. 1100, June, 1961.

Chapter 31

MECHANICAL REALIZATION OF PATTERN RECOGNITION

Paul Metzelaar[*]

IT IS CLEAR THAT LIVING SYSTEMS ARE BETTER SUITED TO THE RECOGNITION of patterns than are present-day machines. A child can distinguish its parents from other people long before it can analyze anything in a logical way. With machines just the reverse is true; they can rapidly and accurately perform arithmetic and logical operations, but they have not yet done comparably well at distinguishing people or other entities in the outside world. In the performance of the great majority of pattern-recognition tasks, machines are many times slower and less versatile than people.

A pattern has been defined as a recognizable and in some way useful structure in a set of signals. Two types of pattern recognition should be distinguished. In the first sense, pattern recognition is the solution of the problem of sorting different events into preassigned classes. Here the possible patterns that may occur are known in advance by the recognizer, and the problem is to choose the ones that are presented for judgment. The second class of problems is that in which the observer is presented with a set of events and is asked to judge what the pattern is that they all have in common. In the latter sense, the use of the term pattern recognition is somewhat misleading, for the pattern is not specified in advance, and therefore it cannot be recognized from the data; it can only be cognized or induced from the data.

Pattern recognition covers a large range of complexity. In visual pattern recognition, one might be asked to identify a faintly visible aircraft or to recognize an entire picture, as in identifying particular elements of a street scene. In acoustic pattern recognition, one might be assigned simple frequency discriminations or such subtle discriminations as are required of an orchestra leader in finding which of his violinists played a false note. There are a great many examples of the simpler forms of pattern recognition by machines. In general, these machines use fixed patterns and have fixed tolerances on these patterns. An example of

[*] Space Technology Laboratories, Los Angeles, Calif.

a simple mechanical pattern recognizer is the teletype receiver and printer. This machine is able to receive and decode the standard teletype signals and convert them into printed text. These machines are familiar, but they are hardly ever thought of as machines that perform a higher function of living systems.

A first example of the more advanced pattern recognizers is MAUDE, the Morse Code Automatic recognizer, developed in 1956–1957 [8]. The sender is a human being with a Morse key and transmitter; the receiver is a MAUDE, a mechanism that receives and decodes the Morse signals and converts them into printed text. Human operators are not exact about the duration of the dot and dash symbols and the letter and word spaces that they transmit. To compensate for this nonuniformity, the machine performs "local averaging" on these durations. If a human sender gradually speeds up or slows down his transmission, MAUDE is able to "follow" and to decode correctly. Whatever the speed of transmission (provided that it does not change too fast), the machine, obeying its built-in rules, will count the longer symbols as dashes and the longest spaces as word spaces. Its adjustable and fixed rules make MAUDE a system that has good recognition capabilities for the signals sent by different operators, although human recognizers can perform better than this machine.

In evaluating the results obtained with MAUDE, it should be remembered that MAUDE consists of just two racks of electronic equipment and a printer. This is a fairly simple mechanism relative to the complexity of the task and relative to the human recognizer with whom its performance is being compared. In the extensive testing done, MAUDE ranked second to human operators in Morse-code reception. Error rates of about 1 per cent are quoted for human receivers. Six per cent character errors still results in a normally readable message in English. If it is assumed for the moment that this rate of error is acceptable, then, in tests on MAUDE, 30 out of 184 messages were unacceptable.

The builders of MAUDE felt that a machine of this type and complexity could increase its performance only slightly by having different rules of the same general complexity built into it. Apparently, decoding machines of much improved performance can be built only at the cost of greatly increased complexity, not only involving ingenious decoding schemes but also making more use of the structure of the language employed and requiring greater understanding of the human decoding process. MAUDE is a very good example of what can be done practically with the use of relatively simple equipment.

One of the most critical steps in the process of designing pattern-recognition machines is the collection of an adequate sample of the types

of inputs that the machine will utilize. This is necessary in the design of any device, but it is particularly important in the design of a pattern recognizer. A careful analysis of the inputs can lead directly to the types of significant features that distinguish the patterns and to the rules that should be applied in using these features. Analysis of inputs becomes a problem of major proportions when it is concerned with such a complex input as human speech. Much work in acoustic pattern recognition has been devoted to the design and construction of speech typewriters, which type out dictation. In addition, a great deal of work has been done in determining the basic nature of speech variables [19, 20], with the result that a good foundation now exists on which to base speech-recognition principles.

Several attempts have been made to develop mechanical speech recognizers. For example, in Fry and Denes's device [5 to 7], the tape-recorded inputs were compared with reference patterns and the phoneme best matching the inputs was selected. In addition, the machine provided some linguistic information in the form of sequential probabilities, which biased the primary acoustic recognition system; "unbiased" outputs were obtained by using the acoustic recognizing system alone. The device supplying the output of the machine was a typewriter that signaled every recognition by typing a particular letter, and therefore it was fairly easy to compare the input sequence with the recognition results. However, the art of automatic speech recognition is still far from realizing a general speech recognizer. In Fry and Denes's experiment, the repertory of sounds with which the recognizer worked was restricted to 12 English sounds. The speech material was a series of 140 English words containing only these 12 sounds. It was found that the machine should be set up to deal with the speech of one speaker and that he had to be instructed to enunciate carefully. With acoustic recognition alone, only 60 per cent of the sounds were recognized correctly. With linguistic biasing the proportion of sounds correctly recognized increased to 72 per cent. The output tended to contain unwanted extra letters, which were counted as errors by the designers. Since each word was made up of several speech sounds, the proportion of words correctly recognized was lower than that of sounds, 44 per cent in the biased condition and 24 per cent in the unbiased.

In a practical phonetic typewriter a self-adjusting or learning system would be necessary unless time-consuming manual adjustments of the machinery from speaker to speaker could be tolerated. In attempting the recognition of the spoken word [1, 2, 18], researchers on speech typewriters have tackled an extremely difficult problem. The search here is still largely that of finding those features which will enable the simplest and most reliable distinction between the spoken words.

As to visual recognition, there have been some practical applications of mechanical pattern recognition developed in the biological area. A scanner with an associated computer has been developed for automatically recognizing possibly cancerous cells among healthy cells displayed under a microscope [26]. Another visual recognition device is a digital-computer program developed for automatic detection of patterns in the electroencephalogram (EEG) [3]. At present the main tool of EEG classification is the eye of the experienced clinician observing the waveforms recorded directly by pen. Preliminary results with the device have indicated that a number of different human subjects and states of the same subject can be distinguished with excellent reliability.

In the recognition of printed characters, the researchers have more stylized inputs to work with. The simplest approach is again that in which specific features are used for distinguishing between the images. Several researchers have found that one of the simplest methods of avoiding the complications due to variations of location, size, and density is to recognize letters by their character strokes [25]. In this technique letters are scanned. The scanner has a light beam much narrower than the width of a letter and sometimes narrower than an individual line element of a letter. The number of intercepts of the light beam, the length of each intercept, and the relative location of intercepts are used as three basic inputs to the recognition logic.

This approach has found practical applications. For instance, a machine has been developed that can read typewritten material from a card and then punch specific items of information so obtained back into the same card [9, 13]. The machine's speed is 150 cards per minute, and it can read through limited amounts of printing imperfections. The development of more comprehensive pattern-recognition capabilities is being actively pursued by large companies concerned with data processing and communication [10, 12, 21]. In research in all but the most simple pattern recognition, digital computers have usually been enlisted to simulate the action of the pattern recognizer. The comparative influence of various input patterns and the comparative efficiency of various recognition schemes can be more readily studied by changing computer programming than by building special-purpose machines.

Why is the realization of mechanical pattern recognition difficult? A versatile pattern-recognition machine must in some sense be able to compete successfully with a human pattern recognizer. This means that machines must be designed that can recognize patterns in spite of changes in one or more of the following factors: (1) overall size; (2) intensity; (3) contrast gradients; (4) position in the field of view, angular orientation; (5) movements; (6) color, frequency translations; (7) inversion, spatial or temporal symmetry; (8) shape, frequency distortions;

(9) addition of noise, addition of irrelevant or partially obscuring objects or sounds.

It is perhaps useful to recall something that George Miller said in 1955:

> Many psychologists have imagined that some memory trace of every object ever perceived in the past is stored somewhere in the nervous system. Then when a new object is presented, its image is correlated with all the stored images, and that stored image which gives the highest correlation with the perception of the present object determines what we call it. However, when we in fact get around to really building such a machine, the inadequacy of such a model quickly becomes apparent [17, p. 111].

To demonstrate this point, let us compute how many possible images of one of the letters of the alphabet, e.g., the letter A, could be projected on a 100 $\times$ 100 element raster. Let us allow sizes, intensities, horizontal and vertical translations, rotation, and letter styles each to assume 1 of 10 possible values. Then there are 1 million different letters A to be considered. If noise or color variations are permitted on the 10,000-element raster, the different images to be considered are numbered in the billions. If every one of the possibilities has to be considered, as would be necessary in the simplest pattern-matching schemes, the pattern-recognition problem becomes a practical impossibility. When the astronomically large number of possibilities is considered, either the computer cannot practically remember a sufficient number of past images or there would not be enough time to perform enough correlations to make such a scheme feasible. One of the most important general conclusions that may be drawn from the recent research in mechanical pattern recognition is that certain preliminary transformations must be performed first in order to reduce the amount of remembering and correlating that must be done for recognition and to make the job practically feasible.

In the present way of designing pattern-recognition machines, the preliminary transformations are usually chosen by the designer himself, who bases them on a careful analysis of the inputs to be expected and pays particular attention to those input characteristics which should serve to resolve the differences between the input events to be recognized. In this sense, the preliminary transformations to be used depend upon the problem. There are some of these transformations which have been more often used than others. In the context of pattern recognition of two-dimensional figures, several investigators have used spatial transformations [10, 14, 15, 23, 27]. The input area is usually divided into a raster of independent elements that are either black or white, and the state of each element is compared with those of its neighbors. One much-

used transformation is the "cleanup" operation, which is done by finding an element that is surrounded by nothing but elements of the opposite state. This method can be used to eliminate small imperfections in the image. Another operation considers the symmetry properties and tests, for example, whether all neighboring 1's are on one side and all 0's on the opposite side. Depending on the number of neighbors considered, this operation can find edges and corners in the input image.

Using a digital computer, Selfridge and Neisser [24] used 28 different preliminary transformations to distinguish 10 different hand-printed letters. The transformations determined such features as the number of intercepts with horizontal and vertical lines, the number and lengths of edges, their ratios, and concavity features. The preliminary transformations were executed for each new pattern to be recognized. The outcome of the transformations was expressed as the degree to which each feature was present and was entered into a table. The result was a table of the probability of finding each transformation outcome for each letter. The letter with the highest total score for all features was the one chosen for reproduction. This scheme, called Pandemonium, is claimed to make only 10 per cent fewer correct identifications than human readers. Also, Pandemonium can itself be used to help establish the a priori probabilities. In this learning mode it is presented with a random sample of the letters to be recognized, and the outcomes computed are associated with the correct letter by the experimenter to form the initial table of feature-weighted probabilities.

Present-day digital computers are set up with arithmetic units that conveniently handle strings of digits. In picture-processing operations, the picture is usually first shredded into a series of strings of digits. To perform the area-processing operations, it is necessary to preserve somehow the relation between the strings that were originally (in the picture) adjacent to each other. A considerable amount of costly bookkeeping is necessary for keeping track of these relations throughout all transformations. Many investigators, notably Unger and Kamnetsky [14, 27], have felt that it would be worthwhile to allow more of the preliminary area transformations, such as cleanup, edging, and cornering, to be done directly at the input to the machine. From this type of approach there may result a set of area-transformation operations that are directly adapted to the two-dimensional pattern-recognition task. Computers will always seem rather inefficient in their operation with these visual patterns until they are built with components more nearly suited to the task.

In the examples mentioned so far, the mechanical recognizer is somewhat inferior in its performance to the human operator. There are other recognizers whose performance is superior to that of human beings. One of these is Hagelbarger's SEER, the Sequence Extrapolating Robot [11].

This machine is able to simulate a well-known penny-matching game. To play against the machine, the player chooses and lights one of two lights. The machine analyzes the pattern of the player's previous choices and chooses one of a second pair of lights to match his most recent choice. If the machine's choice matches the player's choice, the machine wins. This machine is built with fewer than 100 relays. It can learn to distinguish four patterns in the human being's play and to play accordingly. If none of these patterns are detected, the machine plays at random. This simple device wins in about 55 to 60 per cent of the rounds, whereas, if it is played strictly at random, it would win on the average only 50 per cent of the time. The SEER capitalizes on the inability of human beings to play strictly at random and thus win the 50 per cent of the rounds that is allowed him by the theory of this game.

Another machine that can beat the average human opponent is Samuel's checker-playing program for the IBM 704 [22]. This is perhaps the only example of a program that can store a fairly large background of past patterns (of board positions) together with their values. When reported, the machine had stored 53,000 board positions without the need for excessive searching of the magnetic tape on which the positions are stored. Access to the stored data is obtained through a cataloguing system that groups the board positions by such items as the total number of pieces on the board, piece advantage, presence or absence of kings, and the first movements of the pieces about the diagonal axes on the board. The machine qualifies as a good player and can usually beat its own inventor.

In contrast to pattern-recognition processes, the pattern-cognition process may be considered as proceeding in four main phases, observation, classification, analysis, and decision. In the observation phases, the inputs and the similarities between the inputs are observed and the decisions or predictions made. In the classification phase, the inputs are transformed, and the similarities are computed for the events that are to be cognized. The pattern classifications resulting from these operations are examined in the analysis phase. The classifications may be merged or split, depending on whether redundancy or contradictions are discovered; the boundaries between adjacent classifications may be moved. If the success criteria for the recognition process are defined directly in terms of the patterns to be recognized, the device is a pattern-recognition machine. If only general success criteria are given and the machine is given the ability to find the classification yielding the best performance, then it is a pattern-cognition machine. One simple criterion that has been used to determine success is whether the machine can successfully use the patterns found to predict the future of the sequences of data presented to it. In operation the pattern-cognition machine may be thought of as continually cycling through the four phases described.

A control function serves to adjust the parameters of the search so that the machine more rapidly and effectively converges on the solution of the problem.

On pattern cognition, in contrast to pattern recognition, only a small amount of work has been done. Kochen and Gelanter [16] and Foulkes [4] have worked out the problem of predicting the future of sequences of 1's and 0's. This was done by inspecting the past of the binary sequences to find sequences similar to the most recently received digits. If a similarity was found, the assumption was made that history would repeat itself and a prediction was made accordingly. If the prediction is verified by events, the usefulness of the pattern used to make the prediction is demonstrated; if not, use of the pattern may be discontinued. Pattern cognition is usually thought of as being typical of one of the higher functions of living systems.

The pattern recognizers discussed here mark the beginnings of accomplishment in a vast area of research and development. The advances already made allow application of mechanisms to simple recognition tasks such as receiving and decoding signals, reading simple characters, and recognizing simple sounds. For reasons of simplicity and economy, these simple recognition capabilities may be all that can be afforded for many applications. However, there are yet a great many problems to be solved both at the primitive mechanism level and for more complex system applications. Three of the most important of these problems are resolution control, attention control, and control of multiple-recognition mechanisms.

Resolution control refers to the capability for considering inputs both in fine detail and in gross outline. Present machines work at a constant level of detail. A variable resolution capability would be of great assistance in all problems in which the overall size of the image or the required level of detail cannot be foreseen, e.g., the recognition of objects in aerial photographs.

Attention control is the recognizer's ability to direct its attention to a certain part of the input and to ignore other parts within the field of view. Present recognition machines are lacking in the ability to read when several letters are viewed simultaneously or to concentrate on one speaker when several are talking.

The third capability, the control of multiple-recognition mechanisms, is an advanced system problem that at present seems far from a solution. Digital computers of approximately 100 times more capacity than those available today are on the drawing boards. It has been shown that this increased capacity can be either wasted by fruitless bit-by-bit search through an astronomical number of possibilities or profitably used by increasing the organization and sophistication of current recognition mechanisms or procedures. The limitations of present-day machines are

such that only relatively simple recognition schemes can be implemented. Some devices are able to use tables of a priori probabilities of the presence of a feature in a set of characters; others can adjust the amount of filtering done on the input data or use and control the acceptance or rejection thresholds for the decision that a pattern is present. A basic inventory of mechanisms is being accumulated that can perform these functions for specific inputs and for specific purposes. The new generation of machines will make possible the combination of individual mechanisms, with each playing the role for which it is best suited. The efficacy of such combination has already been demonstrated by some of the work in heuristic programming and by Selfridge and Neisser's work on Pandemonium. Still, however, relatively little is known concerning the technique for designing a sufficiently flexible multimechanism recognition machine. This is perhaps the most difficult technical problem.

It seems clear that the flexibility of the recognition machine in choosing the proper mechanism according to the inputs presented will determine whether or not the device can complete a pattern-recognition problem successfully. Only when such flexibility is achieved can it be said that a higher function of living systems has been mechanized.

BIBLIOGRAPHY

1. David, E. E., Jr.: Artificial Auditory Recognition in Telephony, *IBM J. Research Developm.*, vol. 2, no. 4, pp. 294–309, October, 1958.

2. Davis, K. H., R. Biddulph, and S. Balashek: Automatic Recognition of Spoken Digits, *J. Acoust. Soc. Am.*, vol. 24, no. 6, pp. 637–642, November, 1952.

3. Farley, B. G., et al.: Computer Techniques for the Study of Patterns in the Electroencephalogram, paper presented at Symposium on Pattern Recognition, University of Michigan, Ann Arbor, Mich., Oct. 21–23, 1957.

4. Foulkes, J. D.: A Class of Machines Which Determine the Statistical Structure of a Sequence of Characters, *IRE WESCON Conv. Record,* pt. 4, pp. 66–73, 1959.

5. Fry, D. B., and P. Denes: Experiments in Mechanical Speech Recognition, in Colin Cherry (ed.), "Information Theory," pp. 206–212, Butterworth & Co. (Publishers) Ltd., London, 1956.

6.——— and ———: Mechanical Speech Recognition, in Willis Jackson (ed.), "Communication Theory," pp. 426–432, Butterworth & Co. (Publishers) Ltd., London, 1953.

7. ——— and ———: On Presenting the Output of a Mechanical Speech Recognizer, *J. Acoust. Soc. Am.*, vol. 29, no. 3, pp. 364–367, March, 1957.

8. Gold, Bernard: Machine Recognition of Hand-sent Morse Code, *IRE Trans. on Inform. Theory,* vol. IT-5, no. 1, pp. 17–24, March, 1959.

9. Goldstein, G. D. (ed.): Digital Computer Newsletter, *J. Assoc. Computing Mach.*, vol. 4, no. 3, pp. 371–391, July, 1957.

10. Greanias, E. C., et al.: Design of Logic for Recognition of Printed Characters by Simulation, *IBM J. Research Developm.*, vol. 1, no. 1, pp. 8–18, January, 1957.

11. Hagelbarger, D. W.: SEER, A Sequence Extrapolating Robot, *IRE Trans. on Electronic Computers*, vol. EC-5, pp. 1–7, March, 1956.

12. Harmon, L. D.: The Use of Computers in the Simulation of Pattern Recognition, paper presented at Symposium on Pattern Recognition, University of Michigan, Ann Arbor, Mich., Oct. 21–23, 1957. (Not generally available.)

13. Heasley, C. C., Jr.: Some Communication Aspects of Character Sensing Systems, *Proc. WJCC*, pp. 176–180, Mar. 3–5, 1959.

14. Kamnetsky, L. A.: Pattern and Character Recognition Systems—Picture Processing by Nets of Neuron-like Elements, *Proc. WJCC*, pp. 304–309, Mar. 3–5, 1959.

15. Kirsch, R. A., et al.: Experiments in Processing Pictorial Information with a Digital Computer, *Natl. Bur. Standards (U.S.) Rept.* 5713, Dec. 13, 1957.

16. Kochen, M., and E. Gelanter: The Acquisition and Utilization of Information in Problem Solving and Thinking, *Inform. and Control*, vol. 1, no. 3, pp. 267–288, September, 1958.

17. Miller, G. A.: Comments on Session on Learning Machines, *Proc. WJCC*, pp. 110–111, Mar. 1–3, 1955.

18. Olson, H. F., and H. A. Belar: A Phonetic Typewriter, *J. Acoust. Soc. Am.*, vol. 28, no. 2, pp. 1072–1081, November, 1956.

19. Peterson, G. E.: The Information Bearing Elements of Speech, *J. Acoust. Soc. Am.*, vol. 24, no. 6, pp. 629–637, November, 1952.

20. ——— and H. L. Barney: Control Methods Used in a Study of the Vowels, *J. Acoust. Soc. Am.*, vol. 24, no. 2, pp. 175–184, March, 1952.

21. Rochester, N., et al.: Tests on a Cell Assembly Theory of the Action of the Brain, Using a Large Digital Computer, *IRE Trans. on Inform. Theory*, vol. IT-2, no. 3, pp. 80–93, September, 1956.

22. Samuel, A. L.: Some Studies in Machine Learning Using the Game of Checkers, *IBM J. Research Developm.*, vol. 3, pp. 211–229, July, 1959.

23. Selfridge, O. G.: Pattern Recognition and Modern Computers, *Proc. WJCC*, pp. 91–93, Mar. 1–3, 1955.

24. ——— and U. Neisser: Pattern Recognition, *Sci. American*, vol. 203, no. 2, pp. 60–68, August, 1960.

25. Tersoff, A. I.: Automatic Type Size Normalization in High Speed Character Sensing Equipment, *IRE Natl. Conv. Record*, vol. 6, pt. 4, pp. 318–326, 1958.

26. Tolles, W. E., and R. C. Bostrom: Automatic Screening of Cytological Smears for Cancer: The Instrumentation, *Ann. N.Y. Acad. Sci.*, vol. 63, art. 6, pp. 1211–1218, Mar. 30, 1956.

27. Unger, S. H.: A Computer Oriented toward Spatial Problems, *Proc. IRE*, vol. 46, no. 10, pp. 1744–1750, October, 1958.

Chapter 32

IMAGE PROCESSING AND FUNCTIONAL RETINA SYNTHESIS

*E. E. Loebner**

THIS CHAPTER IS DIVIDED INTO THREE SECTIONS. THE FIRST SECTION DEALS with current image-processing methods and is limited to a predominantly descriptive discussion of the various schemes proposed and employed. These schemes emphasize pattern recognition and depend heavily on methods derived from programmed digital computers that have been adapted to handle pictorial input data. The analysis of images into classifiable invariant features has so far been limited to the intuitive insight of the investigators and machine designers. No self-contained theory about the human or the machine discriminability of the various image features exists at this time. The basic data needed for such a theory involve the topic dealt with in the second section of this chapter, the structure and function of neuroretinas. Knowledge about the retina could lead to the discovery of the machine language of the visual apparatus that is part of most biological systems. There is reliable evidence that vertebrate retinas carry out image-processing functions of categorization and sorting by means of spatial, temporal, and hybrid attributes of the retinal images [12, 20, 39, 40]. These images carry coded messages about the visually accessible environment. It is possible to deduce that complex signal handling, involving local directional logic, occurs at synaptic fields adjoining the receptors and separated from the higher functional centers within the brain by a number of additional synaptic stages. The functional image processing attributed to some vertebrate retinas can be also studied in retina analogs. This is the topic of the third section of this chapter. The realizability of this analog by means of solid-state optoelectronic panel systems has been investigated and an attempt made to duplicate the retinal-neuron net topology of three cascaded layers of interconnected active elements. It is believed that the resulting model

* hp associates, Palo Alto, Calif.

This chapter was prepared while the author was employed by RCA Laboratories, Princeton, N.J.

will be useful, not only in the design of active input equipment for automatic data-handling machines, but also in investigations of the mechanisms of vision.

IMAGE PROCESSING

In the broadest sense, "image processing" means a functional transformation of a spatial distribution of some physical property through a mapping procedure by means of a specifiable operation. These transformations are of three types, (1) transformations of the quality or quantity of the radiant flux, (2) transformations of the time variations in the image, and (3) transformations of the space variations or of the mapping of the image [23].

Examples of image processing include contour enhancement, contour outlining, and deblurring, or sharpening, of images as reported by Kovàsznay and Joseph [17]. The image processing proposed by these investigators required a conversion from the space domain to time signals by means of an isotropic scan. All the processing operations were carried out in the time domain and were followed by reconversion into the space domain through the use of an extension of television techniques. The reconstructed pictures could be stretched, warped, and repositioned by translation, rotation, or reflection. A drawback of this procedure is the amount of time and the complexity of the electronic equipment required for picture pickup, processing, and reproduction that handles each small portion of the image at separate times in rapid succession.

The history of time-domain processing of pictorial data by mechanical or electronic means can be traced back to the nineteenth century. The conceptually simple but technologically complicated simultaneous and parallel handling of pictorial information appears to have been proposed first by G. R. Carey in 1875 [6]. He suggested a photochemical pickup for a simultaneous television system. Six years later, Ayrton and Perry described a system using selenium cells [1]. The first demonstration of such a system was probably by Rignoux and Fornier in 1909 [44]. The parallel handling of pictorial information has, however, been abandoned in favor of a time-sequential scanning process. Prominent among the early scanning methods were Nipkow's disks [37]. From these disks evolved the cathode-ray methods that conveniently matched the time-sequential communication technologies of the telegraph, telephone, and radio. The present-day conventional picture pickup and reproduction process is based on a scanning procedure totally unrelated to the picture content. This has been shown by many workers to contribute to a huge redundancy and to lead to information-channel and bandwidth waste.

There have been several attempts to alleviate this situation. For example, the work of Mertz and Gray has led to the accommodation of the color-signal frequencies within the existing bandwidth that was originally allocated to black-and-white television only [30, 31]. Cherry and Gouriet have proposed the elimination of the excessive redundancy in television signals by a variable velocity scan alternating in the horizontal and vertical directions and by the storage of every successive frame [4]. Although other methods have been investigated, the problem of finding a suitable code to replace scanning still exists. Intuitively or *ad hoc* selected parameters, attributes, or features of pictorial data and signals for pictorial-data extraction or filtering have not thus far resulted in worthwhile savings or improvements.

Among the early workers to experiment in pictorial-information processing and to use a general-purpose stored-program digital computer was a group headed by Kirsch [16]. They used 47 per cent of the memory capacity of the computer for storage of a two-tone 3.1×10^4 element picture. The processing carried out was numerical and geometrical. The numerical processing was concerned with numbers relating to counts of separate objects and the area occupied by each object; the geometrical processing was concerned with center-of-gravity location, rectilinear-image translation, image reversal, and contour outlining. Special computer programs were written for each of these operations. Some of these programs led to the discovery of neighborhood-logic operations and the resistance to degrading of noisy images under certain image-processing procedures. Although this work shows that it is possible to use digital computers for image-processing research, it indicates the incompatibility between the problem and the tool intended to solve the problem; highly specialized input and output equipment must be used. The method is awkward in that many seconds are required for carrying out each operation, inasmuch as the picture elements are sequentially processed, and time-consuming programming routines have to be written for each process. Since automatic pictorial-information techniques will be most likely to find applications in alphanumeric-character recognition and specialized picture analysis, it can be argued that only specialized computers would be adequate for the task.

Similar comments may be made concerning the approach taken by Unger [54, 55]. With the help of a conventional large-size stored-program computer, he studied the simulated operation of a novel hypothetical computer oriented toward spatial problems. The problem of matching pictorial input and a stored-program digital computer was solved with a cellular computer design. It consisted of a rectangular arrangement of many identical, simple, small-capacity, general-purpose digital computers that received identical commands from a central control. The

inputs were in the form of an image field registered onto the matrix of computers, to which were added lateral inputs from the four nearest-neighbor computer cells. The topology of the output was symmetrical to that of the input. Each of the computer cells contained a Boolean algebraic unit and several distinguishable memory digits. The total conventional computer components per cell exceeded 10^2. Another way of describing such a computer is in terms of an active memory that is a memory, an algebraic unit, and a shift-register input and output all in one. Unger did not suggest that such a spatially oriented computer might be realizable in the foreseeable future with present components and techniques. Nevertheless the various image-processing operations studied by him via computer simulation are of considerable interest. He investigated space and digital quantization, explicit smoothing processes that included filling of isolated holes and of notches in straight edges, elimination of isolated specks and of bumps on straight edges, and replacement of certain missing corners and characterization of hand-printed alphanumerics through the establishment of a chain of edge-direction sequences. Although Unger's processing operations were carried out simultaneously over the whole image area, a large sequence of commands is needed to carry out even some of the simpler operations; for example, the detection of an alphanumeric character may require over 100 sequential steps.

Another alphanumeric-character detection study employing a conventional large-size computer for simulation is that of Bomba [2]. The four image-processing steps preceding the so-called recognition, i.e., character-decoding, process were (1) denoising, (2) line thinning, (3) character-feature extraction by the generation of 17 extracted pattern fields, and (4) division of each extracted pattern field into 9 subfields. Neighborhood logic was employed throughout the study. The denoising operation used eight-neighbor logic (four at faces and four at corners) processed twice. It employed the filling of empty spots when more than five neighbors were present and erased when fewer than three were present. Line-thinning programs were written for the reduction of lines over 4 but less than 10 elements thick. Each local area for thinning contained, in addition to the processed element, 40 elements divided into 5-element rows in the eight neighborhood directions. The features extracted were straight lines: horizontal, vertical, and slanting lines at 30°, 45°, and 60° to the left and right, four orientations of T intersections, and five orientations of V intersections. The last processing step was numerical. It listed the number of elements in each of the 9 subfields for each of the 17 feature extracts. Bomba recognized the *ad hoc* nature of this method and the absence of a systematic classification of the most criterial features to be tested for. Although his system can classify 34 handwritten block letters on a 5,400-element background, it cannot deal

with serifs and other deviations from the simple letter style. According to Bomba, a system that is not self-organizing and that is intended to classify large alphabets of characters of unspecified styles will probably work with a combination of methods and the addition of feedback. He believes that nonspecific pattern-recognition hardware is something far off in the future, since the complexity of the program required for even some of the most simple preparatory processing operations is fairly large.

Recently a number of current pattern-recognition schemes were evaluated by Kirsch, Minsky, and Neisser [15]. Kirsch stressed that no pattern-recognition scheme is truly universal and that all contain built-in heuristic criteria that optimize certain tasks. He suggested that reference to general-purpose pattern-recognition devices is at least misleading and is most likely to turn out to be false. He also stressed the need for some kind of unified terminology, and possibly theory, that might ease comparison of seemingly unrelated approaches. Minsky stressed that patterns represent classes of figures and that the pattern structure is determined by the means of figure production. The reading process is considered a decision operation that assigns each figure to its prototype, or pattern, class. This process can follow a scheme in which statistical inference, based on evidence supplied by a set of computed image functions, provides the criterion of classification. Such image functions are concerned with noise and distortions that may be threefold, (1) figure-independent and position-independent (additive random noise and therefore removable by correlation techniques), (2) figure-dependent but position-independent (such as smearing), and (3) both figure- and position-dependent (such as size). Minsky also touched on the serial (tree) vs. parallel decision-scheme controversy. He pointed out the heavily weighted malfunctioning of trees in cases of unreliable forks at the top of the tree. Neisser further developed the comparison of serial vs. parallel processing. He pointed out that a tree network is ill suited for learning from experience, since the localization of the error and the retracing of the decision steps in the reciprocal direction are difficult. He stressed the value of sophisticated parallel-processing procedures, which can handle position dependence by processing and shape dependence by selective information attenuators. Contrary to Kirsch, Neisser felt that theory would provide little help and that experimentation should be the approach at this stage of investigation. He based this conclusion on his feeling that such image processes are analogs of higher mental functions in human beings.

Recently the problem of visual pattern recognition has been subjected to a penetrating analysis by Minsky [32]. He divided the problem into three approaches, (1) identification of images by a process of normalization and template matching, (2) the extraction, grouping, and

manipulation of image features and attributes (called property lists), and (3) a complex functional analysis (breakup) of patterns using a spatial descriptive code relating more primitive pattern elements. Minsky follows here the procedure of most pattern processing and recognition research, which neglects nonvisual cues and frames of reference. This appears to be a disturbing omission from a bionics or cybernetics point of view. It could be the case that the axiomatic reliance on all-visual input information represents a stumbling block for meaningful categorization in the image-analysis field. The more fundamental categories, dealing with the three-dimensional-space metric and displacive motion, are derived from intersensory processing [26]. It is they that provide the ordering parameters for the visual input in living organisms. It might be useful to examine the subsumed axiom that the geometrical informational content of pictorial inputs alone is sufficient for the generation and evolution of machine languages that can express geometrical and topological analogs.

Work dealing with image and pattern analysis was described at the 1961 Western Joint Computer Conference. Among the pertinent contributions were those of Frick [7] and Uhr [53]. Frick stated that pattern-recognition problems are peculiarly tantalizing. They appear easy at first. Anyone can generate plausible solutions, schemes, or models, but few significant methods have been developed so far. The four basic problems that still remain are selectivity at the input, extraction of appropriate information, utilization of context, and appropriate decision procedures. All four have to be solved simultaneously. Uhr designed computer programs that generate, evaluate, and adjust image-processing operators. His future work will tend toward a closer analogy with biological systems. Some of the new approaches [52] will make use of logic networks similar to those described in this chapter.

The author's position is, as much as possible, structure-oriented. The structure of pattern-recognition equipment and of figure and form classifications into pattern categories is considered both the key and the lock of the problem. This belief leads one to discard the notion that general-purpose machines are an indispensable analog tool at this stage of the investigation. To simplify the present study, for the beginning at least, those operations which simulate the functioning of higher centers of learning in the brain will not be discussed. A fairly large amount of image processing can be carried out in preprogrammed-decision networks without the necessity of adaptation and learning. If, as Noam Chomsky [5] points out, a four-year-old child is genetically endowed with a decision apparatus for making grammar and syntax discriminations of great complexity, why should not a similar situation prevail in the child's visual apparatus?

Since the author's first description of the synthetic frog retina [21] there have appeared several papers on artificial neural networks, and the literature on this subject is rapidly expanding. This chapter confines itself to the elucidation of a few principles dealing with parallel processing of images in synthetic neural networks and the description of the first model proposed by the author [21]. Over the last year additional work has been carried out [25 to 27]. Currently an advanced version of the synthetic frog retina is under construction.

THE STRUCTURE AND FUNCTION OF THE RETINA

The biological visual system found in vertebrates represents the most highly developed input device yet designed for a pictorial-data-handling computer. Unfortunately, except for the most rudimentary features, its logic design remains unexplored. It has been established that the visual information-processing network operates at three locales that are hierarchically related, the retina, the subcortical geniculate bodies, and the visual cortex. There is considerable evidence that, whatever the processing, a large degree of mapping of the image of the external visual field is preserved at all three anatomic levels.

Both genetically and embryologically the retina represents a part of the brain that had separated and became a highly organized distinct input device. One can hardly approach this field of inquiry without referring to Polyak's "The Vertebrate Visual System" [41], which discusses the essential neurons and their synaptical network topology, i.e., external circuit connections, in the primate retina. A more up-to-date but less encyclopedic discussion can be found in Brindley's recent book [3]. A large number of experiments on the electrical activity of the retina has led to a voluminous amount of literature on the subject, which is only a small portion of the more general topic of electrical activity in the neural networks of animals and man. Not only is this subject extensive, but it is also complex; therefore, it cannot possibly be discussed in its entirety in this chapter. However, full use will be made of the formal analogy between computer elements and neurons as understood by Wiener [56], McCulloch [29], and von Neumann [34]. A neuron, including neurons in the retina, will be considered as an active signal-transmission element that receives one or more input signals and produces an output signal that can be distributed to one or more inputs of other neurons.

A neuron [50] is the structural unit of the animal nervous system. It is composed of a cell body, called the perikaryon, or soma, from which extend a number of conducting processes. One of these processes is an-

atomically different. It is called the axon and can have collateral branches. Its terminal arborizations are called synaptic knobs, or telodendrites. The other processes are anatomically similar to the cell body and are called dendrites. Neurons that lack dendrites are called unipolar. The branching topology of neurons varies greatly throughout the nervous system. It is commonly assumed that the signal input is received by a neuron via its dendrites and on the cell body directly and that the signal output is emitted via the axon. The place of junction of neurons, i.e., where the axonal-end arborizations of one neuron come into contact with the cell body or dendrites of another neuron, is known as the synapse. It is normally assumed that synaptic junctions are strictly unidirectional, permitting signal transmission only from axonal endings of one neuron to the dendrites or body of the other neuron. A neural network can thus be specified by a connectivity diagram or synaptic field of neurons, provided that the functional characteristics of the individual neurons, i.e., their input-output behavior, are known. It therefore becomes possible to describe formally a neuron network as "Boolean junction boxes" connected into a complex network. It has been the belief of a large number of anatomists that the functions carried out by nerve cells and signaled by their axons are specific to their anatomy, i.e., that each different type of neuron carries different sensations. This was first stated by Müller [33]. However, irrespective of a large amount of circumstantial evidence, this correlation has been doubted by prominent psychologists [38]. In anticipation of work described later, it is assumed here that Müller's law is valid. This discussion will therefore adopt this point of view when describing and analyzing the neural network in the retina. Anatomically different neurons will be viewed as representing network elements whose input-output function, and thus the logic of signal processing, is also different. It is stressed at this point that the form and nature of the signal itself are being ignored. The shape of the spikes, their magnitude, and other properties are of no concern, for they act only as information carriers, just as the electrical space charge in a tube, a cloud of minority carriers in a transistor, a configuration of domains of magnetization in a transfluxor, or a photon flux in an optoelectronic unit does not necessarily specify a different logic circuit. On the contrary, network analogs using elements having different signal carriers have been constructed.

Topologically there appear to be at least two types of networks in the retina, a transverse signal-handling network and a lateral signal-handling network. A diagram of the transverse retinal signal network in primates is shown in Figure 32-1. Direct pathways are provided by the transverse retinal signal network for electrical impulses transmitted from the photoreceptors to the brain. The numbers on the left of the figure designate the conventional stratification of the 300-μ-thick multilayered

retina. The layers in strata 5 and 7 are called plexiform layers. They contain the two synaptic fields that couple the primary (strata 2 to 4), secondary (stratum 6), and tertiary (stratum 8) neurons, respectively.

The primary neurons are the well-known biological photoelectric transducers, rods *a* and cones *b*, terminating with round spherules and conical pedicles, respectively, in stratum 5. The secondary neurons *d*, *e*, *f*, *h*, shown in Figure 32-1, are called bipolar neurons. They are also

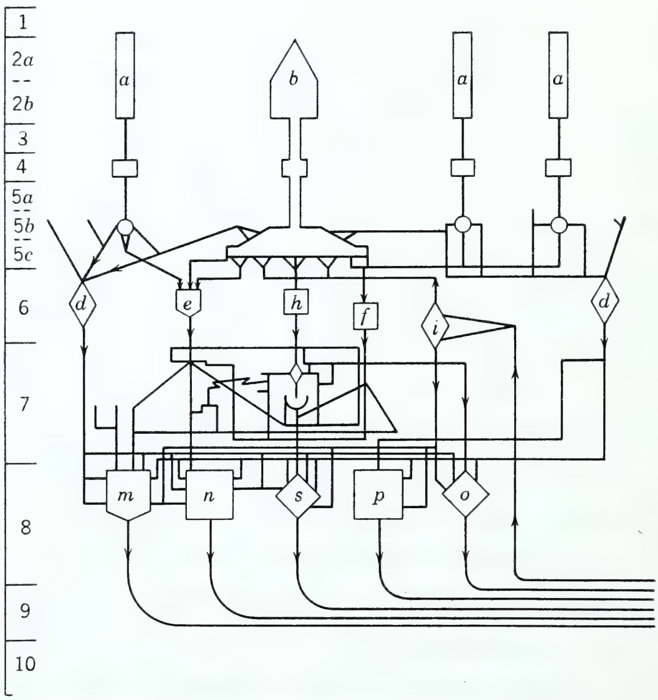


FIG. 32-1 *Diagram of the transverse neuroretinal network in primates.*

known as signal-converging, or centripetal, neurons. These neurons receive their input in the first synaptic field from the photoreceptors and provide an output in the second synaptic field to the tertiary neurons. The tertiary neurons *m*, *n*, *o*, *p*, *s* are also known as ganglion cells. The sum of their output fibers is a bundle called the optic nerve, which terminates in six major layers of the geniculate bodies. From there neural connections called optic radiations carry visual impulses to the visual cortex. Throughout all these layers a functional derivative of the retinal optic field map remains conserved.

Not considered in the following discussion is a third network, which apparently represents a small fraction of all the neural connections to

and within the retina. It contains centrifugal, i.e., signal-diverging, bipolar association neurons *i*. They exhibit connections to the cones, to the ganglions, and directly to the brain. Their function is not clearly understood, but it appears most likely that they provide feedback to the receptors from the brain and ganglion cells. Polyak hypothesizes that the feedback is inhibitory [41]. Such neurons would be capable of changing the functioning of the other neurons and in a sense provide a visual-perception bias at the retinal level. Such effects have been recently proposed by Pritchard [42]. He found that stabilized images fade and regroup in meaningful patterns. Such phenomena could be due to an intervention of higher processing centers at the retina by means of fibers conducting to neurons *i*.

The neuron *d* is known as the mop bipolar variety. It receives its inputs via a spread-out filamentary mop of dendrites that make contact to a compact group of rods and cones. There is partial overlap with the group of rods and cones connected to the neighboring mop bipolar. Contact is also made with other varieties of bipolar and associational neurons. The output of mop bipolars is via axons that extend deep into the ganglion-cell layer and provide two types of synaptic connections, (1) a few telodendric axonal arborizations making contact with ganglion dendrites and (2) the majority of connections through direct adherence to the bodies of up to four ganglion cells.

The flat-top bipolars *e* and the brush bipolars *f* receive their inputs from both cones and rods. They are, however, separated from the mop bipolar inputs at the pedicles. This is presumed to be an indication of functional sorting at the photoreceptor synapses. The outputs of the brush and flat-top bipolars differ more from the outputs of mop bipolars than their respective inputs do. For example, the outputs of the brush and flat-top bipolars terminate in a much higher layer, are much finer, and never make contact with the bodies of the ganglion cells. However, all varieties of diffuse bipolar neurons make contact with all varieties of ganglion cells.

A special bipolar neuron is the midget bipolar *h*. It is presumably found only in human and simian retinas and primarily in the foveal region. It receives its input from one cone only, via a so-called private monosynaptic connection. It provides an output exclusively to dendrites of ganglion cells, whereas it avoids the ganglion cell and amacrine cell bodies. Connections to the midget ganglion cells *s* are also strictly private, i.e., monosynaptic, so that in these cases a pure signal transmission, without processing, can occur from the receptor to the brain.

There are a number of types of tertiary neurons *m, n, o, p, s* called ganglion cells. In simian and human retinas there are at least five varieties. All except the monosynaptic midget type *s* are called diffuse,

since their numerous dendrites overlap. They all have only one output axon or optic-nerve fiber. The various ganglion neurons are as follows:

1. The umbrella, or parasol, ganglion *m* terminates in two separate narrow subzones within stratum 7 but shows considerable lateral extent increasing to giant dimension as the retinal periphery is approached. Its dendrites make contact with numerous brush, flat-top *e*, *f*, and midget bipolars *h*. There are few body synapses with mop bipolars *d*. The function of this neuron appears to be one of summation over more or less large areas of the retina.

2. The shrub ganglion *n* has relatively few dendrites that make contact with a group of midget bipolars *h* and with brush and flat-top bipolars *e*, *f*. The mop bipolar *d* telodendrons make contact with its body. Their synaptic network suggests a somewhat more restricted summation process than that of parasols.

3. The small diffuse ganglion *o* has a few synapses in the sixth layer, with a small number of bipolars of all varieties.

4. The garland ganglion *p* has few dendrites that reach large distances in the seventh layer. The garland ganglion dendrites appear to serve a summation function covering relatively large areas of the photoreceptor field.

5. The midget ganglions *s*, which form a private line, have already been discussed.

A diagram of the lateral retinal signal network is presented in Figure 32-2. This network provides indirect pathways for electrical impulses that accompany the direct photoreceptor-to-brain transmission. It connects neighboring neurons, which have been designated by primed letters on the diagram. The lateral signal processing is handled by two varieties of secondary neurons.

First there is the horizontal cell *c*, which presumably receives inputs only from pedicles of cones *b* and provides an output to spherules of rods *a′* and pedicles of cones *b′* that are some distance away and are therefore not necessarily activated coincidentally with rods *a* and cones *b*. The horizontal cells thus provide lateral connections at the first synaptic field. Second there is the "amacrine" cell *l*, discovered by S. Ramon y Cajal [43], which seems to perform a function in the second synaptic field that is analogous to the function of the horizontal cell *c* in the first synaptic field. It shows specific interconnections among several varieties of ganglion cells.

The two diagrams show that excitation of rods leads to signal transmission to all transversely associated bipolars (except the midget ones) and ganglions. Excitation of cones leads to signal transmission to all transversely associated bipolars and ganglions and to signals in neighboring rod and cone systems as well.

Why is the retina so complicated? It is the opinion of the majority of investigators that each part of the complex structure serves some definite and specific function of the visual system. The following are some visual qualities that could be associated with certain, at present unidentifiable, structural constituents of the retina: (1) achromatic light sensation (brightness); (2) color vision (hue, saturation, color constancy); (3) acuity (resolving power); (4) space perception (size,

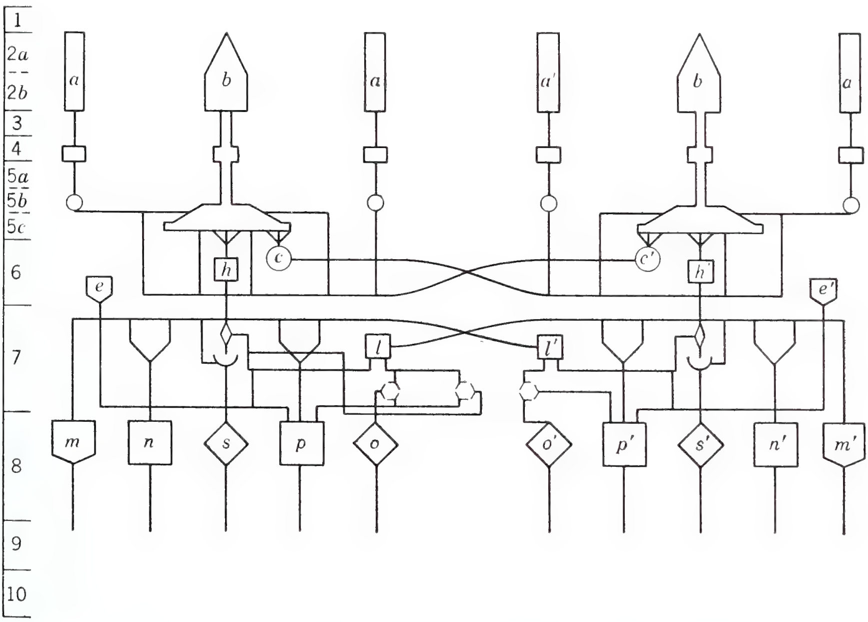


FIG. 32-2 *Diagram of the lateral neuroretinal network in primates.*

interposition, perspective, motion parallax, accommodation); (5) motion perception; and (6) contrast. If the network topology of the retina just described is considered, some estimates can be made about the upper limit of the number of various possible distinguishably different decision pathways through this network. If first the lateral interconnections are neglected but allowance is made for the divergence and convergence of the transverse pathways, there could be up to about three dozen different image-processing functions carried out by this network. The number of possible functions that include lateral cross coupling and feedback is probably an order of magnitude higher.

So far the maze of structures in the retina has been a flexible ground for vision theorists, whose prime preoccupation has been with the solution of the problem of color perception. As an example of the weak

foundation of present-day theory, let it be said that so far the presence of pigment in cones has not been verified and there appears to be no anatomically distinguishable threefold variety of cones. Nevertheless the polyreceptor color theory has many adherents. Even for those who believe that color coding is a synaptic function there are available a great many possibilities as to the specific type of neuron and its location.[1]

Fortunately, in recent years attention has been directed to the study of other retinal functions. Among those who deserve a special mention is Platt [39], who applied electrical-network and amplifier concepts to the discussion of neural networks. He has discussed a number of topics dealing with the retina [40]. His approach is based on the concepts of information processing in a decision network. Platt has shown how the scanning operation of the eye and the geometrical function of self-congruence can give rise to a null detection of visual pattern elements such as straight lines and parallel lines. He has also proposed that pattern perception is equivalent to the finding of addresses in an originally nonaddressed mosaic. It appears, however, that a long learning period is required for such systems.

A criticism of Platt's and similar learning-type approaches, such as Rosenblatt's Perceptron [48], may be based on the obviously nonrandom network found in the biological visual systems. The argument that the complexity of a nonrandom wiring diagram for a system of over 10^9 elements would be almost impossible for chromosomes to specify does not stand up under scrutiny. There are not just two extremes, a system determined completely, to the last fiber, and a completely undetermined system of random connections. There is the possibility of roughly specifying the presence and location of functional neural subnets but leaving the details of the connections to chance without affecting the gross functional features and nature of the system.

Support for this view as well as for the long-suspected but never proved functional specificity of various retinal neuron varieties and their associated nets comes from the work carried out on the frog retina. The frog possesses a number of significant simplifications in its visual apparatus. Its eye does not scan and has no fovea [20].

In spite of its simplified organization and functioning, the frog retina contains many, from the present point of view essential, features described above for the primate retina. The frog retina has two types of receptors, rods and cones, and two types of networks, transverse and lateral. There are several types of bipolar cells. One type spreads out dendrites that connect exclusively with cones; it distributes its axons widely, but in a single substratum. Other bipolar types arborize their

[1] The reader interested in this topic is directed to a most interesting set of papers by S. A. Talbot [51].

axons in several substrata in the second synaptic field. The frog has ½ million ganglion cells. On the basis of shape and size of dendritic trees, there are five types of ganglion cells. Just as in primates, there are two types of secondary neurons that form the lateral network. These are horizontal cells and amacrine cells.

Although the biological function of the frog's vision is much simpler than that of primates, there appears sufficient comparative similarity on the anatomic level of the retina to warrant the hope that many solutions arrived at for the frog will find some, even though possibly transformed, counterpart in higher animals. Others have also confidently assumed that conclusions based on the frog can be transferred to other vertebrates [3]. Thus for over 80 years the frog retina has served as one of the most widely studied visual systems. However, it was only in 1938 that Hartline solved the difficult problem of recording from single nerve fibers [11]. He analyzed his results as follows: There are three kinds of fibers: 20 per cent of the fibers give a burst of impulses when light is switched ON and continue to discharge as long as light is continued; 50 per cent give bursts of electrical impulses only when light is switched ON and when it is switched OFF; 30 per cent respond to OFF light only. This interpretation was based on a purely temporal analysis of the signal. Further work by Hartline [8, 10, 43, 44] and Barlow [25] showed that the different types of fibers of single ganglion cells respond to different sizes of receptive fields, i.e., numbers of receptors centered mapwise on the ganglion cell, and that ON-OFF fibers are sensitive to motion of a light spot across the frog retina. Similar results have been obtained for the ganglion cells of the light-adapted cat [18].

In 1959 Lettvin et al. [20] succeeded in reclassifying the frog optic-nerve fibers into four classes, which are presented in Table 32-1. Each class was assumed to carry out a complex spatiotemporal operation

TABLE 32-1 *The Four Sorted-out Perceptions and Their Transmitters in the Frog*

Perceptions	Transmitters				
	Ganglion axons	Speed of conduction, m/sec	Receptive field, deg	Memory	Hartline's designation
Sustained contrast...	Unmyelinated	0.2–0.5	2	1 min	On
Net convexity.......	Unmyelinated	0.2–0.5	6	<1/10 sec	
Moving edge........	Myelinated	2	12	....	On-off
Net dimming........	Myelinated	10–20	15	Long	Off

SOURCE: J. Y. Lettvin et al., What the Frog's Eye Tells the Frog's Brain, *Proc. IRE*, vol. 47, pp. 1940–1951, 1959.

on the input image. The frog retina thus was reported to perform a sorting operation in which signals about detecting sustained contrast, net convexity, moving edges, and net dimming, originating from receptive fields of different sizes, are carried with different speeds and different delays along different kinds of fibers. These findings represent a significant step forward in the understanding of the visual process. The first two functions are time-independent. The last is the only neighborhood-independent function. Prior to the work of Lettvin et al. the proposed lateral interactions in the retina were primarily simple additive inhibitory or facilitory functions [9, 12]. In a more recent publication Lettvin et al. [19] expanded the classification to five types, which correspond to the five types of ganglion cells in the frog retina. Also the spatial properties of types I and III were reported to be interchanged. All these changes were incorporated into later designs [25 to 27] of analog retinas. However, they are not material to the principal discussion of devising an analog to simulate the retinal behavior.

Lettvin et al. also showed that the analogy of the feature-extraction process to some of the preprocessing operations that have been discussed in this paper is inescapable. Knowledge about what the frog does with this information in the higher centers would probably lead to the solution of how the frog responds to flies, i.e., the correlation part of the decision process. From present information it can be deduced that little learning is involved on the retinal level of processing. Both the addresses and the programming are predetermined in the network.

The attempt to equate the functioning of the eye and that of electronic equipment is not new. A number of contributions have been made in this area within the last 15 years. For example, Rose explored a unified approach to the eye and synthetic image-handling devices [47], and Zworykin and Flory investigated electronic aids for the blind [57]. A mechanism based on variable gain has been proposed as a model for dark adaptation by Rose, who also demonstrated the quantum limitations of the visual process [46]. Schade constructed a color-TV model of the eye [49]. The functional analogy of nonscanning solid-state image-transducing techniques to the processing in the retina has been pointed out by Loebner [24]. The present chapter attempts to exploit the majority of functional features found in the retina of the frog [20] in order to design a synthetic retina that would approximate that of the frog.

FUNCTIONAL RETINA SYNTHESIS

In the attempt to design a synthetic retina the use of optoelectronics is attractive for several reasons. Optoelectronics provides a natural

language for image processing from a variety of aspects; that is, the input is optical and does not require conversion or processing. The output, left in the form of an optical image, provides easy access for a human investigator. The nonreciprocity of optoelectronic nets resembles closely that found in neuron nets. The natural divergence of optical paths from a light source and the falloff of intensity with distance approximate the arborization of telodendrons, and the illumination of one photoconductor by many light sources is analogous to the bushing out of many dendrites from the neuron-cell body. The wireless light-beam connections provide a fairly good substitute for the complex neural network texture found in the retina at the synaptic fields. There are additional advantages of having light for the signal carrier. It eliminates the necessity of wire-tapping the studied system and its individual stages for diagnostic or observational purposes. Solid-state optoelectronic networks can be microminiaturized inexpensively.

Solid-state optoelectronic active devices have been used since 1955 [22]. Originally thought of as light-amplifying elements, these devices were soon recognized as interesting arithmetic and logic elements as well. The Boolean OR function is obtained by mutual paralleling of photoconductors, whereas the AND function follows from placing them in series. In a way these functions correspond to the neural functions of facilitation and summation. The logic negation, realized through a photoconductor shunt of the electroluminor, corresponds to the neural inhibition. The possibility of light-controlled threshold logic in units following the McCulloch neuron rules [28] has been demonstrated [24].

In addition to serving as logic elements, optoelectronic cells have been shown to be suitable for functioning in large-area image-transducing equipment such as monochrome [14] and two-color light-amplifier panels [36] and image-converting and image-storing panels of various constructions [13]. The synthesis of the logic-, i.e., neuronlike, and image-handling capabilities of optoelectronic elements had been foreseen [24]. Recent work carried out by Hook and Giaimo has confirmed these expectations [45].

Complex image-processing operations have been carried out on optical feedback and cross-feed monochrome panels by using the effects of unidirectional current facilitation and inhibition of photoconductivity and its light-threshold-dependent fatigue. Details of the operation and performance of these panels and others capable of image reversal, image quantization, line thinning, form outlining, and motion detection are discussed elsewhere [45]. Optoelectronic panels are especially well suited for such operations because of their relative simplicity of fabrication and operation. Their operation represents a close electronic analog of a number of electrochemical photographic processes such as image

reversal and solarization, as well as visual psychophysical phenomena such as contrast enhancement and reversed afterimages.

A design of a synthetic retina by the use of several optoelectronic mosaic panels is discussed below. The signal-flow diagram of the panels is shown in Figure 32-3. With the exception of the motion-detecting

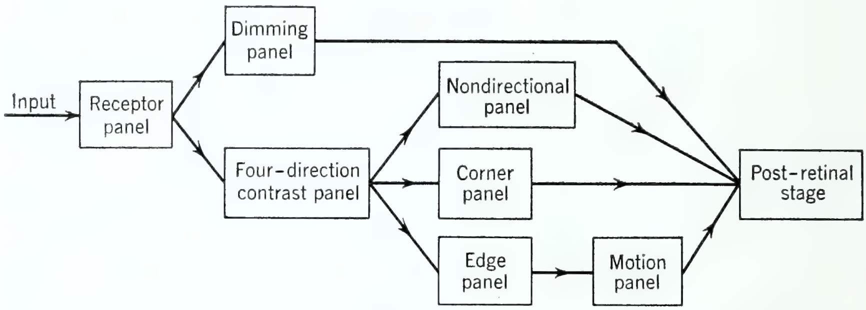


FIG. 32-3 *Optoelectronic panel system and signal-flow diagram of a functional analog of the frog's neuroretina.*

panel,[2] the system has been limited to two processing stages, which correspond to the two synaptic fields in the retina. The various specific panels are shown separately in the diagram, although only the technological difficulty of building them into an interlaced composite panel prevented their being presented in an integrated monolayer fashion. Functionally it does not matter whether the image is split into subimages in a global or in a local manner.

The input, or primary, image is split into two images. The first image impinges on a "dimming" panel, which consists of an array of dimming cells to be discussed in detail below. Since no neighborhood logic is involved in it, this processing step most closely resembles the direct, or "private," neural lines of the simian retina. The function that is carried out in this panel is a combination of image reversal and image storage. The output of the dimming panel provides one of the four outputs of the whole system. In frog retinas an areal summation takes place in this as well as the other feature-extracting systems. This model retina does not provide for areal summation, whereas the later designs do [25 to 27]. The second portion of the split primary image impinges on the so-called multidirectional contrast-detecting panel, which is discussed in considerable detail below because it best exemplifies the powerful processing capabilities of optoelectronic panels. This panel is a Boolean panel made up of

[2] The functional reclassification reported by Lettvin obviates the need for a separate motion-detecting panel.

an array of inhibitor elements arranged in a systematic fashion among the primary receptor array. The output of this multidirectional contrast, or inhibitor, panel is split three ways and is fed into three tertiary summation panels. It is interesting to note that, just as in the biological retina, the first synaptic field of the synthetic retina has a centrifugal tendency, whereas the second synaptic field exhibits a centripetal tendency. It is shown below how three different kinds of sums performed locally on the multiple contrast panel can lead to a threefold image-feature extraction, isotropic or nondirectional contrast, corners, and edges. Whereas the multidirectional contrast panel corresponds to the bipolar cells in the biological retina, the three kinds of summation panel correspond to the ganglion cells. In Figure 32-3 the output of the summation "edge" panel feeds into a motion-detecting panel. Although several structures and circuits have been considered and some have been built and operated [45], a detailed discussion of this specialized motion-detecting panel would unduly prolong this chapter and therefore the panel's structure and operation will not be described.

Let us now discuss the other five panels in detail. The most novel of these is the multidirectional panel, shown in Figure 32-4. The design follows that of previously proposed picture-processing panels based on the incorporation of logic components in the spaces between the image elements [24]. In this panel the intrapanel directionality is obtained by a color code. This form of coding by color is quite arbitrary and does not refer in any way to color vision. Any other space-differentiating form of coding would be equally good. A quadruple grid of color-transmission filters is formed as shown on the diagram. A broad spectrum-image input is incident through the compartments containing the hatched image-element squares. Since the letters R, Y, G, B indicate red, yellow, green, and blue transmission filters, the corresponding color light becomes sorted left, above, right, and below, respectively. Each of the compartments between the input-image elements contains two inhibitor circuits that combine a facilitory with an inhibitory function. The circuits in compartments between two vertical neighbors operate as follows: The circuit on the left can receive a facilitory blue input from above and/or an inhibitory yellow input from below and will provide a red output to the left only when the blue is on and the yellow is off. The circuit on the right is facilitated yellow from below and inhibited blue from above and supplies its green output to the right only when the yellow is on and the blue is off. Similarly the compartments between the horizontal neighbors supply blue below or yellow above if they are facilitated from the left or right, respectively, and not inhibited from the other sides. Each quadrant of inputs has 1 output in its center. Although there are 16 possible input combinations for each quadrant, this system pro-

vides only 7 possible outputs, four individual colors, two pairs of colors, and no output at all. Each individual-color output responds to any one of 3 possible inputs. Thus, for instance, the blue output indicates the presence of an upper left input alone, or in combination with a lower left input, or in combination with both lower, i.e., left and right, inputs. The absence of an output indicates either the absence or the presence

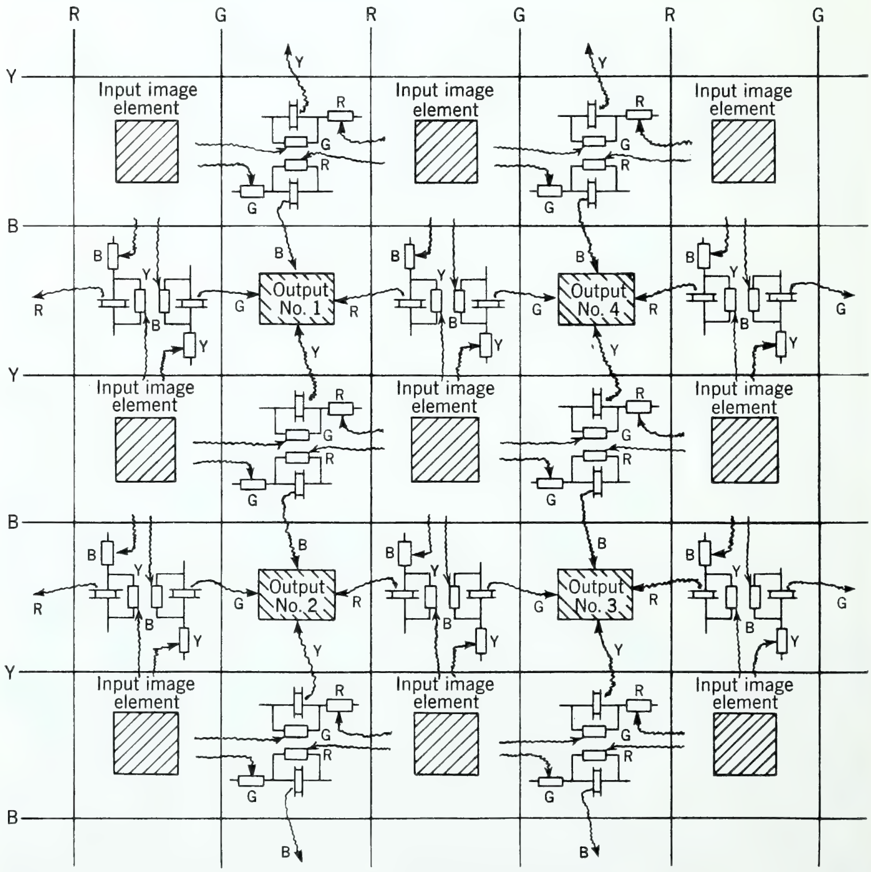


FIG. 32-4 *Diagram of a multidirectional-contrast optoelectronic image-processing panel.*

of all inputs. Only the coincidences of red-green and yellow-blue are unique. They stand for the diagonal input lines slanting with slopes $+1$ and -1, respectively. About 30 per cent of the information is thus lost from the primary image input; additional processing such as selective summation can result in valuable image-feature extraction.

Before discussing the synthetic retina, let us digress somewhat to examine further the idiosyncrasies of this inhibitor panel. For simplicity,

it is possible to change the color code at the center of each image-element quadruple back into a position code. Such a transformation is used in Figure 32-5. The inhibitor panel shows thus an interesting homomorphism, a mapping of many into few. Such a mapping, if suitably selective, underlies the actions of all living things. A closer examination of the input-output relation indicates an alternative possible description of the mapping process. Consider a quadruplet of image elements as a

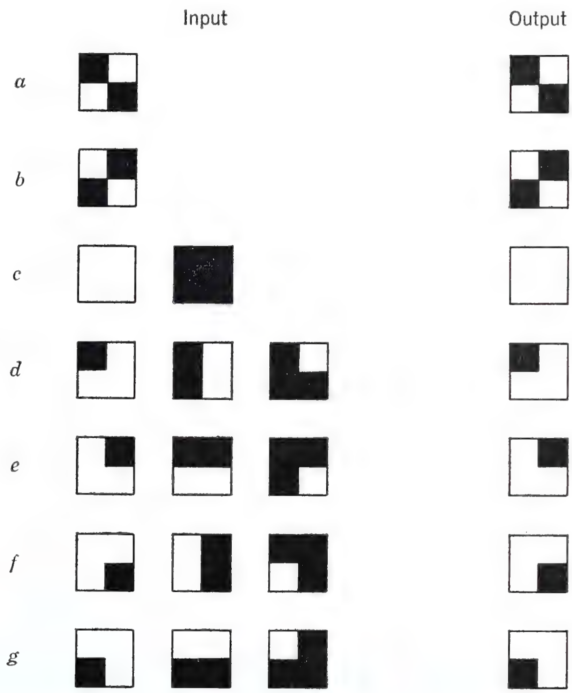


FIG. 32-5 *Homomorphic input-output mapping of the multidirectional-contrast panel.*

four-member ring. Consider now an operation carried out while one proceeds from member to member in a counterclockwise direction. The operation will consist of registering all those ON elements which are preceded by an OFF element. Thus the two patterns *a* and *b* at the top of Figure 32-5 map into themselves, since each ON (black) element is preceded by an OFF (white) element in the ring. For a continuous sequence of all ON elements (as in *d, e, f,* or *g* followed by *c* all ON), only the first one is registered. This results in one 2:1 and four 3:1 mappings, as can readily be seen by inspection of Figure 32-5. There are undoubtedly other useful panels and transformations. No further

digression into the interesting field of Boolean panels will be made here. It is sufficient to say that the inhibitor panel discussed here is a member of a larger class whose properties might be interesting to investigate.

Let us now consider the structure and circuitry of the remaining four panels of the synthetic retina. Figure 32-6 shows the circuit of the

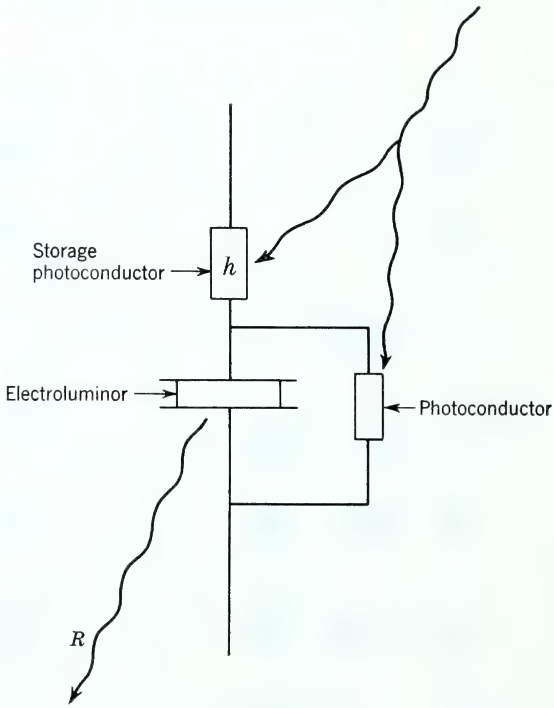


FIG. 32-6 *Optoelectronic circuit of a single cell in the "dimming" panel.*

elements in the dimming panel. This panel performs no local operations; spatially, therefore, it represents an isomorphous mapping. It contains a hysteretic, i.e., storage-type, photoconductor [35] in series with an electroluminor, which is shunted by an ordinary photoconductor. Such an element provides an output pulse only when illumination is removed and for as long as the hysteretic storage lasts. The nondirectional contrast circuit in Figure 32-7 has an output in response to any color input from the differentiating contrast panel of Figure 32-4. It therefore sums over all possible contrast detectors. It is nondirectional and fails to give an output only if all the inputs in the quadrant are either ON or OFF. The edge cell is shown in Figure 32-8. It responds to the coincidence of summation of a predetermined number of mutual vertical yellow or blue

neighbors or mutual horizontal green or red neighbors in the preceding contrast panel. It detects the presence of vertical and horizontal edges. The corner cell is shown in Figure 32-9. The form of a corner has been substituted for that of convexity used by Lettvin et al. [20] in order to conform to the local operations chosen for this discussion. The circuit shown in

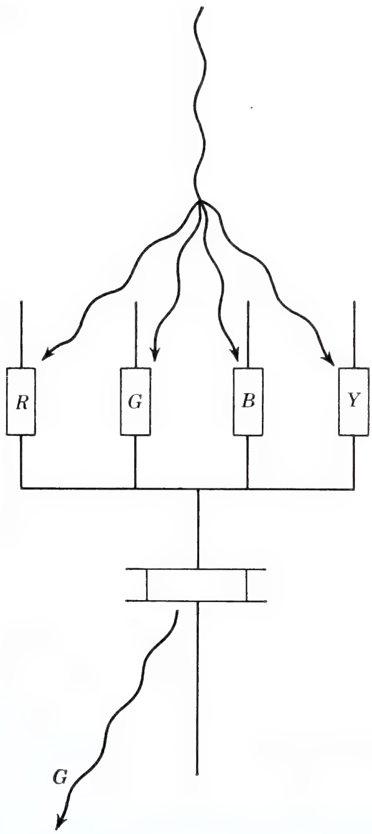


FIG. 32-7 *Optoelectronic circuit of a single cell in the "nondirectional contrast" panel.*

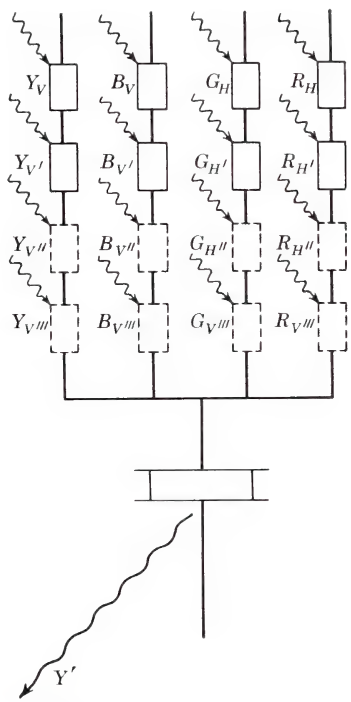


FIG. 32-8 *Optoelectronic circuit of a single cell in the "edge" panel.*

Figure 32-9 sums over four neighboring elemental quadruplets, i.e., over nine input-image elements (the full matrix of Figure 32-4). It carries out the Boolean decision of whether or not the input element, with which the output is associated, is a corner element. It thus detects eight rotational positions of 90° corners and eight rotational positions of 135° corners.

Figure 32-3 shows that the system of panels, which comprises the synthetic retina, has a number of interesting properties. It has a single input image and four output images, which, if in different colors, could be spatially superimposed onto each other without information loss. This data-handling organization represents a functional analog of the frog retina described by Lettvin et al. [20].[3] However, it does have a few shortcomings. Only nearest-neighbor interactions have been included, and the observed variations in the frog's receptive field have been ignored. In addition, the complex time dependence of the convexity

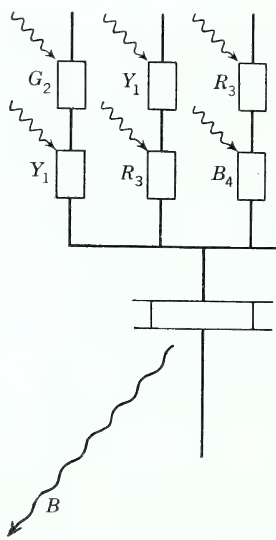


FIG. 32-9 *Optoelectronic circuit of a single cell in the "corner" panel.*

signal has been eliminated. These shortcomings have been overcome in the later models.

Further implications in the structure of the proposed analog retina can be foreseen if threshold logic and sequential rather than parallel feature extraction are employed. However, more complex higher-level processing equipment, incorporating storage and a number of auxiliary panels, would be required.

The model presented in this chapter contains a number of features that are reminiscent of biological systems. Information sorting at synaptic fields of the kind discussed by Polyak [41] and decision networks of the kind postulated by Platt [40] are readily recognized. There are private

[3] In the frog the differentiation between feature extracts occurs in the subcortex, where each function is mapped into a distinct neural layer.

monosynaptic connections at the top of the diagram and polysynaptic ones at the bottom. It appears that two synaptic fields are sufficient for most of the required processing; therefore, there is evidence that the model might resemble living retinas, not only functionally, but also in some structural aspects.[4] This is consistent with the requirement of a cell diversification in the model, which is also reminiscent of the large diversification of neurons found in the retina. Although no specific correspondence is suggested at this time, it is hoped that an expanded catalogue of synthetic retina cells might aid the neurophysiologist in his future work by providing him with topological and circuit models that he can test experimentally. The information-handling specialist might find such proposed synthetic image-processing equipment, which has been intentionally patterned after biological systems, a novel research tool for the study of image-processing techniques and theory. The computer designer might find that the current generation of blind and data-spoon-fed computers could use visual equipment of its own. A number of computer functions, currently handled in the interior of computers, might be advantageously separated and incorporated into their sensory input equipment, just as the retina separated from the brain during the biological evolution. Finally, the programmer might find that homomorphic mapping of parasensory inputs into functionally reorganized memory locations would significantly speed up the job of supplying computers with an agreeable diet of input data.

BIBLIOGRAPHY

1. Ayrton, W. E., and J. Perry: La Vision par l'électricité, *Les Mondes*, vol. 54, pp. 518–521, 1881.

2. Bomba, J. S.: Alphanumeric Character Recognition Using Local Operations, *Proc. EJCC*, vol. 16, pp. 218–224, 1959.

3. Brindley, G. S.: "Physiology of the Retina and the Visual Pathway," Edward Arnold & Co., London, 1960.

4. Cherry, E. C., and G. G. Gouriet: Some Possibilities for the Compression of Television Signals by Recoding, *Proc. Inst. Elec. Eng.*, vol. 100, pt. 3, pp. 9–18, 1953.

5. Chomsky, Noam: "Syntactic Structures," Humanities Press, Inc., New York, 1957.

6. DuMoncel, T.: "Lumière électrique," vol. 2, pp. 267–398, Bureaux du Journal, Paris, 1880.

7. Frick, F. C.: Pattern Recognition, paper presented at Western Joint Computer Conference, Los Angeles, Calif., May, 1961.

[4] This is indeed so, since the latest work of Lettvin et al. [19] has shown that summation is carried out in the second synaptic field.

8. Hartline, H. K.: The Effects of Spatial Summation in the Retina on the Excitation of the Fibers of the Optic Nerve, *Am. J. Physiol.*, vol. 130, pp. 700–711, 1940.

9. ———: The Modification of Sensory Information by Neural Interaction in the Eye and Its Relation to Vision, *IRE Trans. on Med. Electronics*, vol. 6, pp. 84–85, 1959.

10. ———: The Receptive Fields of Optic Nerve Fibers, *Am. J. Physiol.*, vol. 130, pp. 690–699, 1940.

11. ———: The Response of Single Optic Nerve Fibers of the Vertebrate Eye to Illumination of the Retina, *Am. J. Physiol.*, vol. 121, pp. 400–415, 1938.

12. ——— and F. Ratliff: Inhibitory Interaction of Receptor Units in the Eye of the Limulus, *J. Gen. Physiol.*, vol. 40, pp. 357–376, 1957.

13. Hook, H. O.: Optical Feedback Type Storage Light Intensifiers, *RCA Rev.*, vol. 20, pp. 744–752, 1959.

14. Kazan, B., and F. H. Nicoll: An Electroluminescent Light-amplifying Picture Panel, *Proc. IRE*, vol. 43, pp. 1888–1897, 1955.

15. Kirsch, R. A., M. Minsky, and U. Neisser: Discussion of Problems in Pattern Recognition, *Proc. EJCC*, vol. 16, pp. 233–238, 1959.

16. ——— et al.: Experiments in Processing Pictorial Information with a Digital Computer, *Proc. EJCC*, pp. 221–229, 1957.

17. Kovàsznay, L. S. G., and H. M. Joseph: Image Processing, *Proc. IRE*, vol. 43, pp. 560–570, 1955.

18. Kuffler, S. N.: Discharge Patterns and Functional Organization of the Mammalian Retina, *J. Neurophysiol.*, vol. 16, pp. 37–68, 1953.

19. Lettvin, J. Y., et al.: Two Remarks on the Visual System of the Frog, in Walter A. Rosenblith (ed.), "Sensory Communication," pp. 757–776, Massachusetts Institute of Technology Press, Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1961.

20. ——— et al.: What the Frog's Eye Tells the Frog's Brain, *Proc. IRE*, vol. 47, pp. 1940–1951, 1959.

21. Loebner, E. E.: Image Processing and Functional Retina Synthesis, in Bionics Symposium, *USAF WADD Tech. Rept.* 60-600, pp. 309–338, September, 1960.

22. ———: Optoelectronic Devices and Networks, *Proc. IRE*, vol. 43, pp. 1897–1906, 1955.

23. ———: Recent Developments in Solid State Image Transducers, in M. Desirant and J. Michels (eds.), "Solid State Physics in Electronics and Telecommunications," vol. 4, "Magnetic and Optical Properties," pt. 2, pp. 762–775, Academic Press Inc., New York, 1960.

24. ———: Solid State Optoelectronics, *RCA Rev.*, vol. 20, pp. 715–743, 1959.

25. ——— and T. Henneberger: Image Processing in Synthetic Visual Organs, paper presented at International Biophysics Congress, Stockholm, Sweden, August, 1961.

26. ——— and ———: Investigation of Image Analysis Techniques, *RCA Interim Eng. Rept.* 1, Contract AF 33(616)-7734, June, 1961.

27. ——— and ———: Modelling Visual Organs and Perceptions, paper presented at Second Bionics Symposium, Ithaca, N.Y., September, 1961.

28. McCulloch, W. S.: Agathe Tyche of Nervous Nets—The Lucky Reckoners, *Trans. Symposium on Mechanization Thought Processes,* pp. 613–633, 1958.

29. ———: Background, in Bionics Symposium, *USAF WADD Tech. Rept.* 60-600, pp. 51–54, September, 1960.

30. Mertz, P.: Television—The Scanning Promise, *Proc. IRE,* vol. 29, pp. 529–537, 1941.

31. ——— and F. Gray: A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telegraphy and Television, *Bell System Tech. J.,* vol. 13, pp. 464–515, 1934.

32. Minsky, M.: Steps toward Artificial Intelligence, *Proc. IRE,* vol. 49, pp. 8–30, 1961.

33. Müller, Johannes: "Contributions to Comparative Physiology of the Sense of Vision," Cnobloch, Leipzig, 1826.

34. Neumann, J. von: The General and Logical Theory of Automata, in J. R. Newman, "The World of Mathematics," vol. 4, pp. 2070–2098, Simon and Schuster, Inc., New York, 1956.

35. Nicoll, F. H.: Hysteresis Effect in Cadmium Selenide and Its Use in a Solid State Image Storage Device, *RCA Rev.,* vol. 19, pp. 77–85, 1958.

36. ——— and A. Sussman: A Two-color Input, Two-color Output Image Intensifier Panel, *Proc. IRE,* vol. 48, pp. 1842–1846, 1960.

37. Nipkow, P.: German patent 30,105, Jan. 6, 1884.

38. Osgood, C. E.: "Method and Theory in Experimental Psychology," p. 35, Oxford University Press, New York, 1953.

39. Platt, J. R.: Amplification Aspects of Biological Response and Mental Activities, *Am. Scientist,* vol. 44, pp. 180–197, 1956.

40. ———: Functional Geometry and the Determination of Pattern in Mosaic Receptors, in H. P. Yockey et al. (eds.), "Symposium on Information Theory in Biology," p. 371, Pergamon Press, Inc., New York, 1958.

41. Polyak, Stephen: "The Vertebrate Visual System," University of Chicago Press, Chicago, 1957.

42. Pritchard, R. M.: Stabilized Images on the Retina, *Sci. American,* vol. 205, pp. 72–78, July, 1961.

43. Ramon y Cajal, S.: "The Retina of Vertebrates," I. F. Bergmann, Wiesbaden, 1894.

44. Ramberg, E. G.: Personal communication, 1961.

45. Research on Optoelectronic Computer Components, Final Progress Report, Contract AF 19(604)-4959, U.S. Air Force Cambridge Research Center, Air Research and Development Command, Nov. 15, 1960.

46. Rose, A.: The Sensitivity of the Human Eye on an Absolute Scale, *J. Opt. Soc. Am.,* vol. 38, pp. 196–208, 1948.

47. ———: A Unified Approach to the Performance of Photographic Film, Television Pickup Tubes and the Human Eye, *J. Soc. Motion Picture Eng.,* vol. 47, pp. 273–294, 1946.

48. Rosenblatt, F.: The Perceptron, *Cornell Aeronaut. Lab. Rept.* VG-1196-G-1, January, 1958.

49. Schade, O. H., Jr.: Optical and Photoelectric Analogue of the Eye, *J. Opt. Soc. Am.*, vol. 46, pp. 721–739, 1956.

50. Strong, O. S., and A. Elwyn: "Human Neuroanatomy," The Williams & Wilkins Company, Baltimore, 1943.

51. Talbot, S. A.: Retinal Color Mechanisms, *J. Opt. Soc. Am.*, vol. 41, pp. 859, 918, 1951.

52. Uhr, L.: Personal communication, November, 1961.

53. ——— and C. Vossler: A Pattern Recognition Program That Generates, Evaluates and Adjusts Its Own Operators, *Proc. WJCC*, pp. 555–570, 1961.

54. Unger, S. H.: A Computer Oriented toward Spatial Problems, *Proc. IRE*, vol. 46, pp. 1744–1760, 1958.

55. ———: Pattern Detection and Recognition, *Proc. IRE*, vol. 47, pp. 1737–1752, 1959.

56. Wiener, N.: "Cybernetics," John Wiley & Sons, Inc., New York, 1948.

57. Zworykin, V. K., and L. E. Flory: An Electronic Recording Aid for the Blind, *Proc. Am. Phil. Soc.*, vol. 91, pp. 139–142, 1947.

Part H

HUMAN FACTORS METHODS AND PROCEDURES

Chapter 33

PRODUCT AND DESIGN EVALUATION THROUGH THE MULTIPLE FORCED-CHOICE RANKING OF SUBJECTIVE FEELINGS

Edward Bennett[*]

THE INDUSTRIAL OR PRODUCT DESIGNER DESIGNS FOR HUMAN USE. THEREFORE, his success hinges upon public acceptance and satisfaction, for his product must meet both the rational and the irrational needs of his market. As a result, he is keenly interested in knowing what are, or are likely to be, the reactions of people to his design. In order to anticipate and understand these reactions, he must first know what people think about his design. If this insight can be determined through dependable data obtained from the use of scientific methods, all the better.

Customarily, the users of the designer's product have free choice; they accept or reject his product in competition with other products in their market place. They also compare his product with an ideal that exists exclusively within their own imaginations. There are a number of forces influencing the way in which a person deals with the products confronting him. One force is exerted by the variety of competing products that might reasonably satisfy his needs. Another force derives from the many competing needs that he must sort in terms of their relative merits. Since he cannot possibly satisfy all his needs at the same time, he must decide which of the available products can satisfy his most important needs at the moment. He also has to reconcile various competing attitudes toward each product before he can elect to use it or reject it. Finally, a person's attitude toward the other users, or potential users, of the product must be taken into account.

These are but a few of the factors that have to be recognized in

* The MITRE Corporation, Bedford, Mass.

The investigations reported in this chapter were made while the author was associated with the Department of Psychology and the Bio-Mechanics Laboratory of Tufts University, Medford, Mass.

dealing with the sociopsychological experience that we think of as product impact. To a large extent the key element in understanding this impact is the mental activity generated by the product. The actual behavior of a person in dealing with a product is a final event in a chain of prior activities some of which can be called thinking. A person has a pattern of thoughts about the product itself. He also thinks about himself and how he will feel if he does or does not use the product. He thinks about other products in competition with the one in question. These thoughts are largely at an unconscious level, but his unconscious thought encompasses a rich variety of economic, political, social, and esthetic factors.

To a considerable extent, product acceptance is tied closely to the general feeling of pleasure or displeasure associated with the product. High acceptance suggests satisfaction, gratification, and a general feeling of well-being in connection with the product. Low acceptance suggests the opposite. Each person experiences product satisfaction in his own way, since he has his own special constellation of needs to be satisfied. Nevertheless, once this pattern of idiosyncratic feelings has been assessed, the patterns for people in a group can be summed in order to reflect some overall group pattern of product satisfaction and acceptance. The revealed dimensions of product acceptance can then be studied at any of a number of levels: for the individual, his particular group, the general community, or an entire culture. All that is required is that the individual thought patterns appear on the same set of dimensions, no matter how variable the intensities of each of the individual feelings involved.

Three concepts are relevant to an analysis of product acceptance. Stated in question form these concepts are: First, what kind of "personality" does the design or product have? Is this "personality" loaded with pleasant or unpleasant factors? What are these factors? Second, what are the personal reactions of the users, or potential users, to the product? Do the person's attitudes toward himself change with use of the product? Are the users satisfied or dissatisfied with it? What specifically are the high points in their individual reactions? Third, what are the personal reactions that the users, or potential users, think all other users, or potential users, have toward the product? What do the users see as the general acceptability of the product, aside from their own particular feelings?

Considering these three aspects in turn, let us first examine the "personality" of the design or product. It is easily recognized that some products are felt to be "friendly," "cold," "aggressive," "conservative," "fatherly," "casual," and so forth. The use of descriptive adjectives or phrases in this way suggests that a product is thought of in much the

same manner as one might think about a human being. Just as it is possible for a person to describe his impression of the personality of a friend or acquaintance, so too he is able to describe an automobile, a suit of clothing, or a typewriter in the same terms. Knowing something about people's impressions of a product's personality enables us to investigate the origins of the more intense pleasant and unpleasant aspects of this personality impression. It is equally possible to analyze changes in people's impressions of a product's personality as conditions vary.

The second area worth considering is a person's feelings toward himself as a user, or as one who refuses to be a user. It is important if a person develops a more positive attitude toward himself as a result of using a product. It is equally important, from the opposite point of view, if he feels negatively toward himself when he contemplates using the product. Knowledge of such positive or negative feelings may direct the attention of a design analyst to the parts of the product's design that are its emotional assets or liabilities.

The third concept relates to a person's feelings about other people who use, or refuse to use, the product. Although a person can never really know how other people feel, he develops impressions about their feelings and it is these impressions which count for him. If he thinks that a product is giving others the positive feelings that he desires for himself, then he desires the product for himself. The opposite, of course, is equally true; another person's feelings of discontent with a product are usually good enough reasons for not desiring the product.

Successful design analysis usually depends upon being able to gain dependable insights into what the potential and/or actual user is thinking. The key problem, the one that is considered here in some detail, is how to obtain useful and precise insights into such subjective feelings and thinking about industrial designs and the varied products of technology.

METHODS OF ASSESSMENT

In order to obtain information about users' evaluations of products, it is necessary for the people being studied to make judgments about the products. These judgments must be depended upon as evidence of what these people are thinking and of what they feel about the design product.

The more rigorous and quantitative the judgmental task, the greater are the chances of learning something of value about the mental set of the person being questioned. If we were merely to ask people, in a general fashion, to tell us what they think about something, the result

of our inquiry would probably be of little value. At best, we should have diverse opinions that are difficult to group and hence difficult to reduce statistically. On the other hand, if we establish an orderly judgmental procedure, there is a possibility of obtaining meaningful information. Recognizing this inherent need for the precise collection of data on feelings and impressions, one can use psychophysical methods as means for obtaining objective information relevant to subjective experience.

Methods of Rating and Ranking[1]

Briefly, psychophysical methods for obtaining information about a person's concepts fall into two categories, ratings and rankings.

Methods of rating require a person to decide how much of some one quality an object possesses. Additional ratings may assess as many other qualities as needed. Let us assume, for example, that we wish to learn something concerning how people feel about four different designs for a lightweight suit and that we are interested in three qualities, comfort, ease of movement, and pleasantness of texture. Using a method of ratings, we would present a person with one of the suits, a scale of some type to assess comfort, and the instruction to rate this suit by indicating its position on the comfort scale. This would be followed by ratings for ease of movement and pleasantness of texture. In this way 12 ratings would be obtained for the four suits in regard to the three qualities of comfort, ease of movement, and pleasantness of texture.

Methods of ranking, on the other hand, permit a person to compare many objects in terms of one quality. Each object is compared with the other objects on the basis of this single quality. In this way, each object is placed in its appropriate position in relation to the other objects being ranked. Although many things can be ranked, they can be ranked on only one quality at a time. For example, we might ask a person to rank the four suits of the previous example in terms of comfort, ease of movement, and pleasantness of texture. In this way three sets of rankings would be obtained, one for each of the three qualities, comfort, ease, and texture. It would also be possible to have people rank the three qualities for one suit. That is, a person could be asked to decide whether a suit was more comfortable than easy to move in, more pleasant in texture than comfortable, and so forth. Such a request would generate four sets of rankings, one for each of the four suits. The first ranking method tells us something about which suit is better than another, in terms of a given quality. From the second method,

[1] For additional information on the psychophysical methods, see Guilford [11], Johannsen [13], and Stevens [15].

we learn which of the three qualities needs first attention in redesigning each of the suits. In this way the people being studied could be instructed to evaluate a product in a large number of ways, depending upon the demands of the problem at hand.

Polydiagnostic Method of Multiple Forced-choice Rankings[2]

The polydiagnostic method of multiple forced-choice rankings combines some of the characteristics of each of the previously noted procedures. The method is generally as follows: The object or objects to be studied are specified. Then one or more sets of qualities are devised to include all those qualities assumed to characterize in significant ways the object in question. The person whose subjective feelings are to be assessed is presented with each set of qualities in turn and uses them to judge a given object. Just as there may be more than one set of qualities used, there may be more than one object judged during each evaluation. In advance of the actual investigation, the sets of qualities are prepared according to simple rules, as discussed in detail in Bennett [4]. In general, the number of items, let us say n, in a set must be divisible by h, the number of steps or degrees of sensitivity desired in the final rating scale. The quotient of this division, q, is the number of items that the subject chooses at one time in his effort to describe the object under investigation. From the set of n qualities before him, the person under study is asked to select those q which he feels meet some standard established for him. Then from the remaining $n - q$ qualities he again chooses q according to the same standard. From the then remaining $n - 2q$ he once again selects q. Then he selects q from $n - 3q$, and so on, until only the last q qualities remain unchosen. Each of the qualities chosen in the first set of q is given a cardinal score of $q - 1$. The next chosen q qualities are scored $q - 2$. This continues until the final, unchosen q qualities are scored $q - q$, or zero.

To date only sets of 15 qualities and five points of rating have been used, thus requiring the subject to choose 3 qualities at a time. The first 3 qualities chosen as most descriptive of the thing in question are scored +4, those in the second group of three +3, the next three +2, the next +1, and the remaining 3 terms are scored 0. For ease of administration, the set of qualities is often presented to the subject as shown in Figure 33-1. The subject chooses 3 of the 15 qualities and

[2] See Bennett [4] for complete details on the structure, uses, and interpretation of the polydiagnostic procedure, including standardized sets of English descriptors and Flemish, French, German, and Spanish translations. Introductory information can be found in Bennett [3, 5], Bennett and Cohen [6], and Bennett, Cohen, and Kemler [7].

covers them as in step 1 of Figure 33-2. He then chooses 3 more from the remaining 12 (step 2), and so forth, until he has covered all but 3 (step 4). His selections are then scored as described above.

Any other arrangement might prove acceptable, depending upon the ability of the subject to make fine or coarse judgments and to select

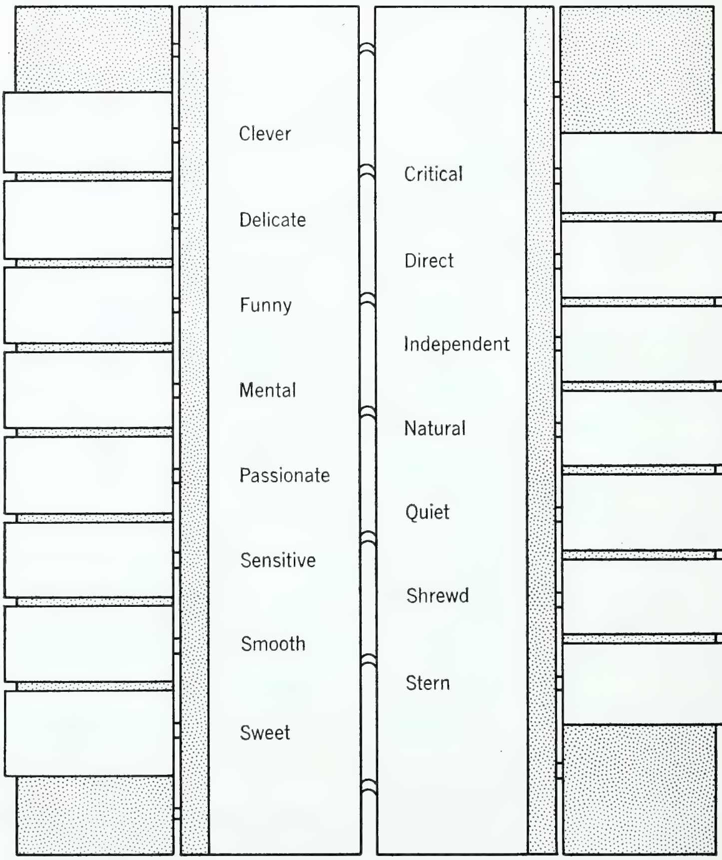


FIG. 33-1 *The administrative procedure for obtaining multiple forced-choice judgments.*

few or many qualities at one time. Currently, however, no data exist for other arrangements.

STUDY 1: AIR AND RAIL TRAVEL

This study was concerned with assessing some of the feelings that people have about air travel and, for comparison, about the more

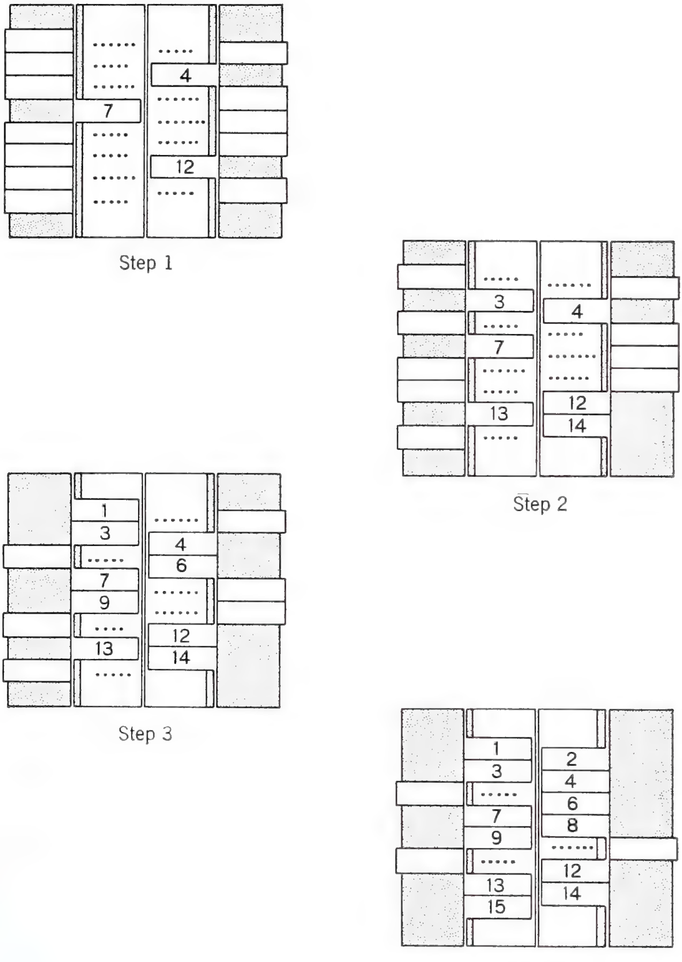


FIG. 33-2 *Figure 33-1 continued.*

traditional experience of rail travel.[3] There were 47 subjects, preselected as having traveled at least 200 miles by either airplane or train within the previous year.

The standardized polydiagnostic procedure for multiple forced-choice rankings was used. Five sets of 15 terms each were selected from the standardized index. These lists are presented in Table 33-1. The subject was presented with the first set of terms and asked, "How do you think people in general feel when they travel by airplane? From this

[3] This study was done in cooperation with D. K. Bennett nee Kemler and B. T. Levin and is described in detail in Reference 9.

set of words choose the three which you think most describe the way people in general feel when they travel by airplane." The subject made his choice, and the examiner crossed out the 3 that he selected, leaving 12 terms. The administration continued, "Again choose the three words which you think now most describe the way people in general feel when they travel by airplane." Again 3 words were selected and crossed out.

TABLE 33-1 *Polydiagnostic Association Sets*

*A**	*B**	*C**	*D**	*E**
Brave	Calm	Careful	Angry	Anxious
Cheerful	Clean	Curious	Ashamed	Bitter
Decent	Exact	Faithful	Cheap	Cruel
Firm	Frank	Friendly	Evil	False
Gallant	Generous	Gentle	Fearful	Fierce
Graceful	Grateful	Honest	Foolish	Guilty
Intelligent	Kind	Loyal	Hasty	Helpless
Modest	Moral	Neat	Jealous	Lazy
Original	Patient	Peaceful	Loud	Lying
Powerful	Practical	Prompt	Nervous	Reckless
Proper	Quick	Reasonable	Rough	Rude
Religious	Responsible	Serious	Savage	Shallow
Sincere	Smart	Social	Silly	Slow
Spiritual	Strong	Tender	Vain	Weary
Understanding	Willing	Wise	Wicked	Wild

* These sets correspond to the standardized index described in Reference 4. *A* is standard-qualities set 16; *B*, 17; *C*, 18; *D*, 19; and *E*, 20.

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

In this manner the subject progressed through the five sets of 15 terms each. Each of the 15 items within a set was assigned a score on the basis of when it was chosen. A term was scored +4 if it was chosen within the first choice, +3 if it was chosen second, +2 if it was chosen third, +1 if it was chosen fourth, or 0 if it was one of the three words left over.

The same procedure was used for each subject in respect to rail transportation, except that in the instructions the word "train" was substituted for the word "airplane." Both the word sets and the instructions were alternated in order of presentation. The words within each set, however, were left in their standardized arrangement.

Results

Two scores for each subject for each of the 75 characteristics were obtained. One score was associated with the airplane, the second with

TABLE 33-2 *Emotional Associations with Traveling by Airplane in Order of Intensity*

Characteristic	Mean intensity of association
Significantly high intensity associations:*	
Anxious	3.58
Nervous	3.45
Fearful	3.36
Cheerful	3.30
Helpless	3.26
Curious	3.13
Brave	2.98
Reckless	2.97
Friendly	2.94
Hasty	2.83
Intelligent	2.75
Weary	2.74
Calm	2.74
Prompt	2.64
Practical	2.62
Silly	2.57
Reasonable	2.53
Powerful	2.51
Significantly low intensity associations:*	
Savage	1.45
Slow	1.45
Gentle	1.38
Guilty	1.36
Modest	1.32
Ashamed	1.28
Generous	1.23
Faithful	1.23
Cheap	1.04
Cruel	1.02
Jealous	0.94
Lying	0.92
Spiritual	0.91
Loyal	0.89
Evil	0.89
Religious	0.87
Wicked	0.68
Moral	0.66
Tender	0.64

* One per cent level of confidence.

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

TABLE 33-3 *Emotional Associations with Traveling by Railroad in Order of Intensity*

Characteristic	Mean intensity of association
Significantly high intensity associations:*	
Cheerful	3.40
Weary	3.34
Anxious	3.32
Friendly	3.19
Calm	3.06
Nervous	3.02
Slow	2.98
Proper	2.94
Lazy	2.91
Silly	2.91
Patient	2.89
Careful	2.85
Practical	2.79
Modest	2.68
Responsible	2.62
Sincere	2.62
Helpless	2.62
Reasonable	2.57
Hasty	2.57
Foolish	2.55
Firm	2.53
Significantly low intensity associations:*	
Powerful	1.47
Strong	1.45
Quick	1.45
Clean	1.45
Graceful	1.39
Jealous	1.38
Guilty	1.23
Faithful	1.21
Tender	1.15
Lying	1.11
Fierce	1.06
Original	1.02
Loyal	1.00
Savage	0.98
Wicked	0.96
Moral	0.89
Cruel	0.81
Evil	0.72
Spiritual	0.65
Religious	0.53

* One per cent level of confidence.

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

the train. Mean scores were calculated for each of the 75 characteristics for airplanes and for trains. Theoretically, 4.00 is the highest possible mean and reflects maximum closeness of association of the term with the transportation medium. A mean of 0 reflects maximum distance of association between the term and the mode of transportation. A mean greater than 2.47 or less than 1.53 is significantly different from that expected on the basis of chance at the 1 per cent level of confidence. If the 1 per cent level of confidence is used to establish cutting lines,

TABLE 33-4 *Popularity of Association of Various Feelings with Traveling by Airplane*

Characteristic	Percentage of respondents who chose the characteristic as a first choice to associate with air travel
Anxious	77
Nervous	68
Helpless	66
Fearful	62
Cheerful	49
Curious	47
Reckless	47
Friendly	47
Brave	45
Powerful	45
Hasty	40
Intelligent	40
Practical	40
Prompt	30
Quick	30

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

the 75 association scores can be divided into three groups, those of significantly high intensity, those not significantly different from chance expectancy, and those of significantly low intensity. A breakdown excluding the terms that were not significantly different from chance expectancy is presented for the airplane in Table 33-2 and for the train in Table 33-3.

Another way of ordering the data is to count the frequency with which a given term was selected in the first, second, third, fourth, or final clusters of 3 terms from the original 15 and to select those qualities for which there was a plurality of first choice. The resulting set of qualities can be thought of as most popularly related to the respective modes of transportation. Tables 33-4 and 33-5 present the characteristics

TABLE 33-5 *Popularity of Association of Various Feelings with Traveling by Train*

Characteristic	Percentage of respondents who chose the characteristic as a first choice to associate with rail travel
Cheerful	64
Weary	62
Anxious	60
Calm	57
Friendly	55
Silly	47
Nervous	47
Practical	45
Slow	45
Social	43
Proper	40
Patient	38
Lazy	36
Careful	36
Helpless	36
Responsible	34
Hasty	34
Curious	32

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

that met this criterion; a plurality of respondents assigned them to the first-choice position.

A third way of viewing the data is to determine those qualities for which there is a significant difference in intensity between plane and train travel. Table 33-6 lists all the characteristics for which there were statistically significant differences in the mean scores for airplane and train travel. With the train as a rational base line, the feeling qualities that are associated significantly more or significantly less with the airplane can be noted. Only characteristics for which the mean difference was at least two standard errors of the mean are reported. Such significantly large differences cannot be expected more often than about 2 times in 100 on the basis of chance alone.

Conclusions

In any multivariate social research, in which an attempt is made to keep in pace with the complexity of the actual social phenomenon, a few

TABLE 33-6 *Emotional Factors More/Less Intensely Associated with Flying than with Traveling by Railroad*

Characteristic	Mean difference*
Air travel being:	
Less slow	1.53
Less modest	1.36
Less cheap	1.11
Less proper	1.11
More fearful	1.07
More powerful	1.04
Less patient	1.02
More quick	0.98
More strong	0.95
More reckless	0.86
More brave	0.86
Less sincere	0.79
More original	0.77
More fierce	0.73
Less careful	0.72
More curious	0.70
Less understanding	0.70
Less decent	0.70
More clean	0.67
More helpless	0.64
Less weary	0.60
More graceful	0.59
More smart	0.58
Less kind	0.56
Less lazy	0.55
Less tender	0.51

* Statistically significant at the 2 per cent level of confidence.

SOURCE: E. M. Bennett, D. K. Kemler, and B. T. Levin, Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

rational generalizations often do injustice to the facts. In view of this limitation, it is felt that the pattern of raw data that was obtained outlines a large variety of factors that are only tentatively open to understanding at the present time.

Most strongly evidenced is the fact that air transportation is associated with fear. The terms "anxious," "nervous," and "fearful" have the highest intensity, as indicated in Table 33-2, and are most popularly related to air travel, as indicated in Table 33-4. The source of such fear may be partially evidenced by the high intensity and popular scoring of the term "helpless." The loss of control and the feeling of personal

incapacity that we commonly think of as helplessness are a fairly rudimentary threat. During the process of flying one is blindly dependent upon forces beyond one's control and understanding. These forces, capable of destruction as well as transportation, certainly enhance anxiety when it is evident that one is unable to influence them.

The second most noticeable pattern of feelings related to air transportation is that centered around having to act in the face of obvious danger. There are both high intensity and popular scoring of the terms "reckless," "brave," and "powerful." This constellation of feelings is probably related to the one of fear and helplessness. Undoubtedly, flying in the face of fear and feeling helpless would encourage feelings of reckless bravery.

The speed of air transportation is reflected in the high scoring of the terms "hasty" and "prompt" and the popular scoring of "hasty," "prompt," and "quick." Other positive feelings are seen in the high intensity and high popularity scores for "cheerful," "curious," and "friendly." In a more intellectual context, air transportation is associated with feeling "intelligent," "practical," and "reasonable." It is likely that these positive feelings are the various qualities encouraging air transportation in the face of personal anxiety.

Rail travel is related to a constellation of feelings that are rather adult and responsible in nature. High scores appear for terms such as "proper," "careful," "practical," "modest," "responsible," "sincere," and "reasonable." In this respect, social acceptability and propriety appear to be associated with this means of transportation.

The relative slowness of the train as a means of transportation is apparently reflected in the high intensity and popular scoring of the terms "weary," "slow," "lazy," and "patient."

Still another rail constellation appears to revolve around pleasurable gregariousness. "Cheerful" and "friendly" are both significantly intense and popular associations. "Social" is popularly scored high.

In addition, rail travel appears to have some quality of ridiculousness associated with it. The terms "silly" and "foolish" both have high intensity scores; "silly" also appears in the popular category.

Rail travel also reflects feelings of anxiety, as shown by the high intensity and popular scoring of "anxious," "nervous," and "helpless."

In the preceding discussion, the absolute mean score values for air and rail travel were considered separately. The conclusions may be checked through an analysis of the significant differences in mean scores for these two modes of transportation as indicated in Table 33-6.

Speed appears to produce the largest difference, as seen in the scoring of air travel as less "slow" and, to a lesser extent, as more "quick," less "weary," and less "lazy."

Following this as a close second is the differential scoring of air travel as less "modest." Here some asocial quality appears more closely related to air than to rail travel. When this judgment is combined with the differential scoring of air travel as less "proper," less "sincere," less "understanding," less "decent," less "kind," and less "tender," there is a further suggestion that air travel has an aura of social impropriety attached to it.

A third cluster appears to be related to aggressiveness. Air travel scores the more "powerful," "strong," "reckless," and "fierce." If there is, in fact, as the findings suggest, a feeling of personal aggressiveness related to air travel, this may be one reason why air travel produces socially improper feelings.

The danger and anxiety factor is also present in this analysis. There is a differential scoring of air travel as more "fearful," more "brave," more "helpless," and less "careful."

STUDY 2: THE TELEPHONE

This study investigated people's subjective feelings about three telephone designs: (1) an ideal telephone that exists only in the mind of the subject; (2) the model 300 telephone, shown in Figure 33-3; (3) the newer model 500 telephone, illustrated in Figure 33-4.[4]

Three sets of 15 descriptive adjectives were used. The three sets were:

> Bold, conservative, definite, different, formal, humble, liberal, mysterious, particular, proud, restless, sharp, simple, solemn, suspicious.
>
> Clever, critical, delicate, direct, funny, independent, mental, natural, passionate, quiet, sensitive, shrewd, smooth, stern, sweet.
>
> Cold, cunning, dependent, distant, hard, innocent, modern, old-fashioned, progressive, remote, severe, shy, soft, strict, youthful.

These three sets of terms had previously been balanced and standardized as sets 13 to 15 of the Polydiagnostic Index [4].

The subject was shown the first of these three sets of terms and was given an unattached model 500 telephone. He was asked to handle the telephone as if he were using it, and he was then instructed: "From this set of terms choose the three which you feel most describe this telephone." Having chosen these 3 terms, he covered them with small covers available for that purpose, leaving 12 items exposed. He was

[4] This study was done in cooperation with D. K. Bennett nee Kemler and J. Forrest, and a preliminary report appeared in Reference 10.

FIG. 33-3 *The model 300 telephone.*

FIG. 33-4 *The model 500 telephone.*

then asked to choose the 3 terms from the remaining 12 that he felt now most described this telephone. He then chose 3 from 9, then 3 from 6, leaving 3. In the same manner the subject was asked to report on the second set of 15 terms, and then on the third set.

He was asked to do exactly the same with the model 300 telephone. He was also asked to express his feelings in the same manner about the ideal or perfect telephone.

The feelings of 50 subjects, 25 males and 25 females, varying in ages from sixteen through seventy-four, all of whom were common users of the telephone for both business and pleasure, were studied. Instructions to judge the model 500, model 300, and ideal telephones, respectively, were alternated.

Results

The subjective concept of an ideal telephone is reflected in Figure 33-5 in terms of mean scores ranked in order of magnitude. The vertical dashed lines indicate the points above or below which a mean is significantly different from the chance mean of $+2.00$ at the 95 per cent level of confidence. In a similar manner the qualities for the two actual telephones are ranked in Figures 33-6 and 33-7.

When an analysis of variance is performed comparing all three telephones for each quality separately, there is evidence that 24 of these 45 qualities are scored significantly different as a function of the type of telephone at the 95 per cent level of confidence. There is no such evidence for the remaining 21 terms. Figure 33-8 indicates the significant terms, with those which seem related grouped together.

Conclusions

The data permit a variety of possible interpretations. Depending upon the needs of the person considering the findings, there are clues for design and product acceptance, or there is simply a picture of how an inanimate object takes on measurable anthropomorphic qualities at a subjective level.

The comparison of both the old and the new with the ideal points up relatively successful or unsuccessful design changes in the personality traits of the telephone. An improvement in design is evidenced where the new telephone is rated closer to the ideal than is the old. Design failure is evidenced where the old telephone is judged as closer to the ideal than is the new. Little difference between the new and the old, both designs being far from the ideal, indicates that there is a need to improve the old telephone and that the new telephone has been

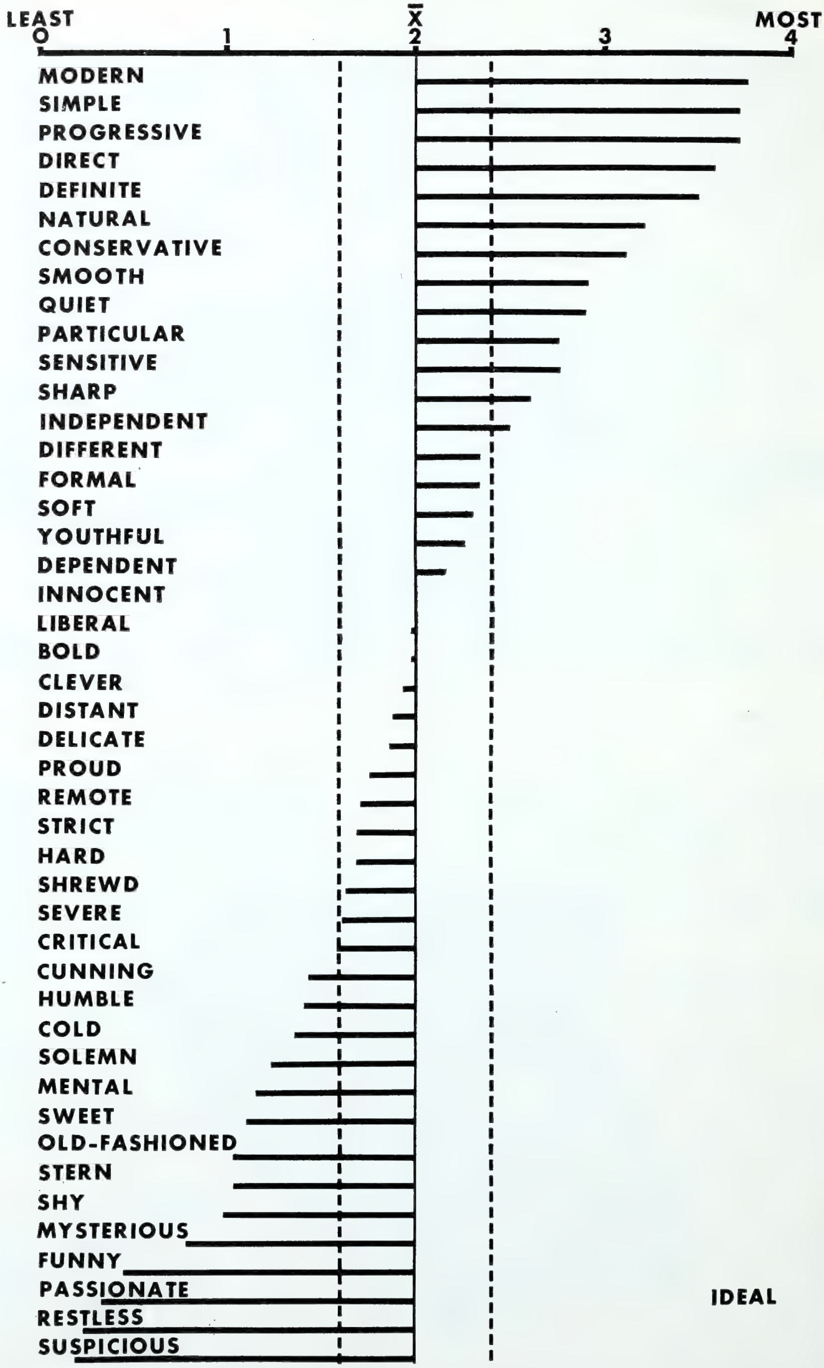


FIG. 33-5 *Subjective judgments concerning the nature of an ideal telephone. The vertical dashed lines indicate points above or below which mean scores are significantly (LC = .95) large or small.*

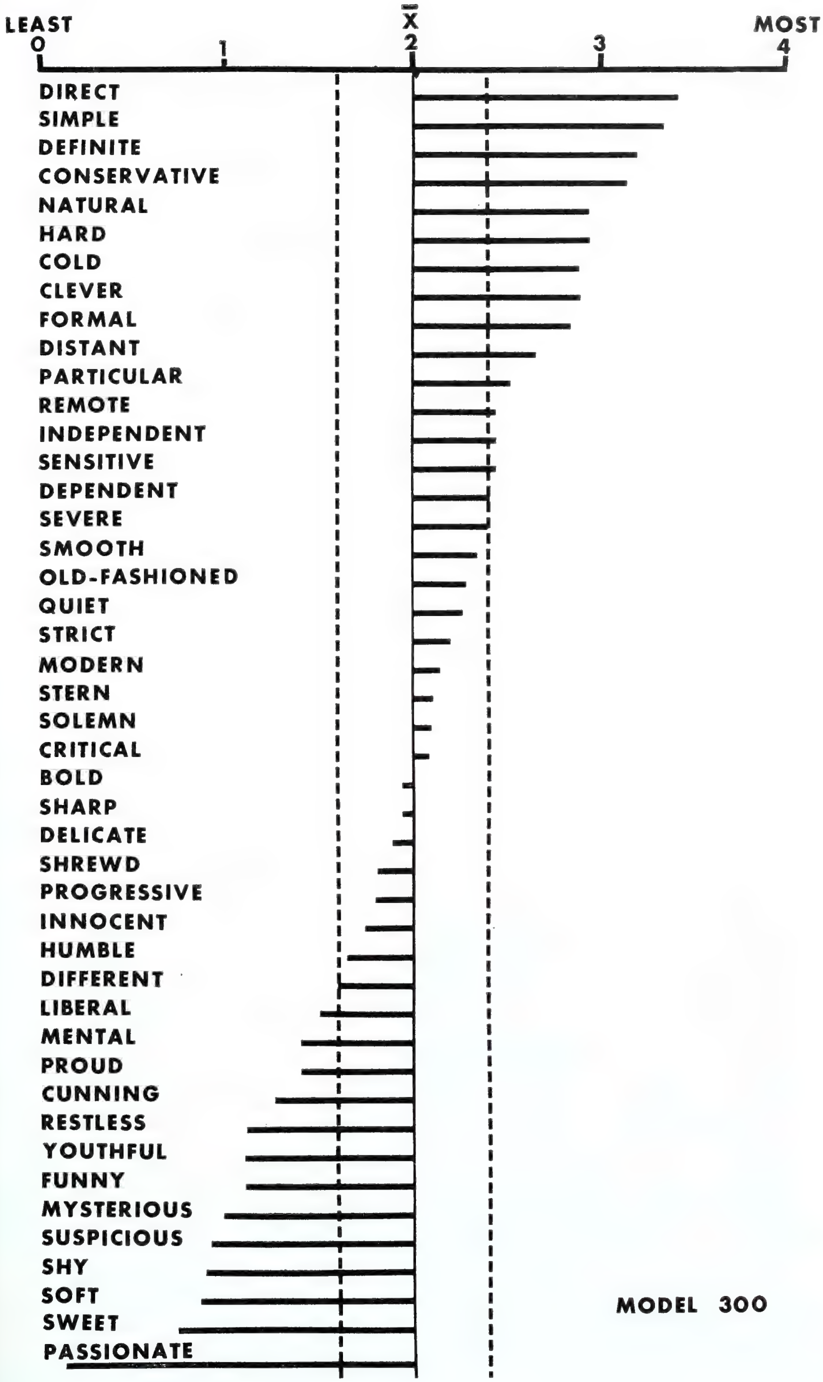


FIG. 33-6 *Subjective judgments concerning the nature of the model 300 telephone. The vertical dashed lines indicate points above or below which mean scores are significantly (LC = .95) large or small.*

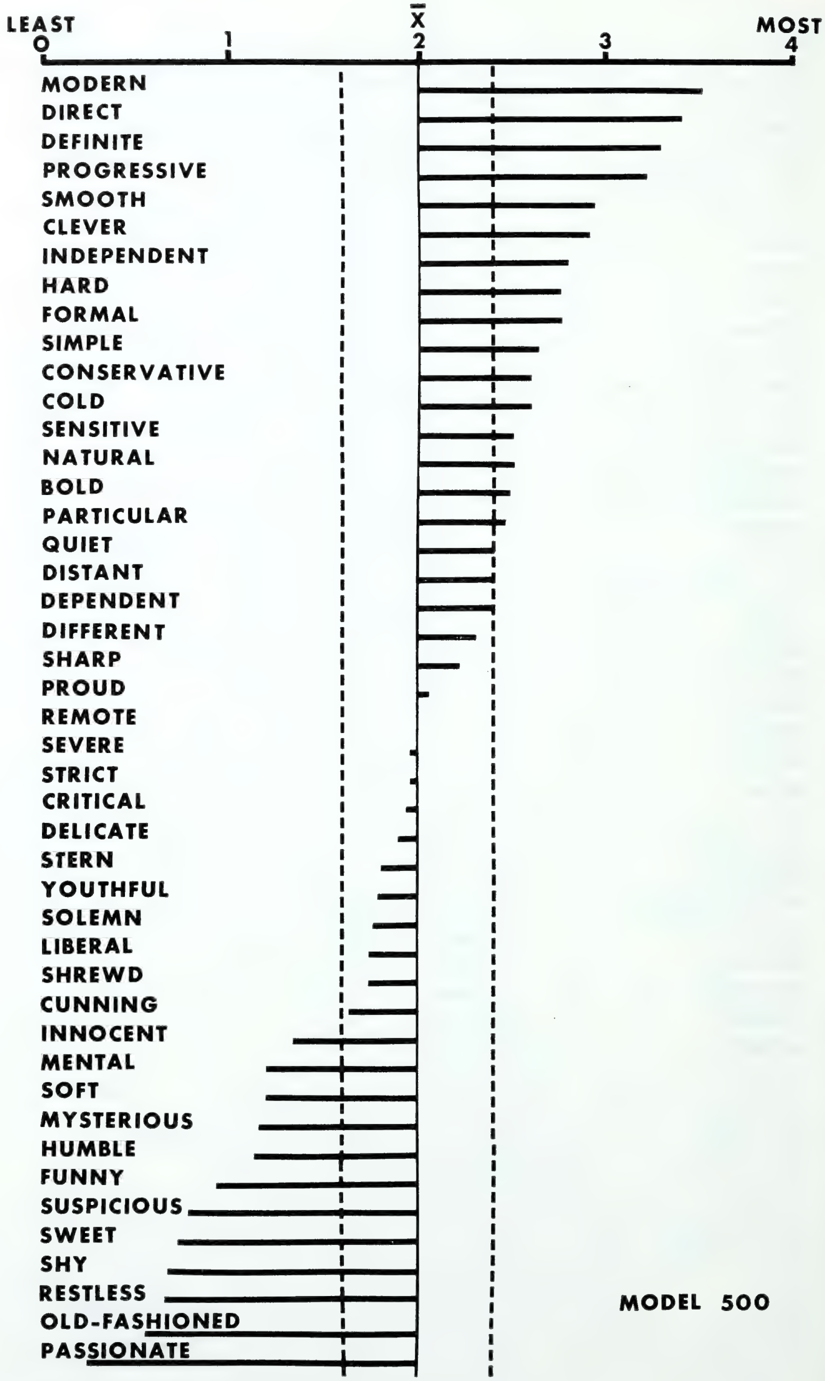


FIG. 33-7 *Subjective judgments concerning the nature of the model 500 telephone. The vertical dashed lines indicate points above or below which mean scores are significantly (LC = .95) large or small.*

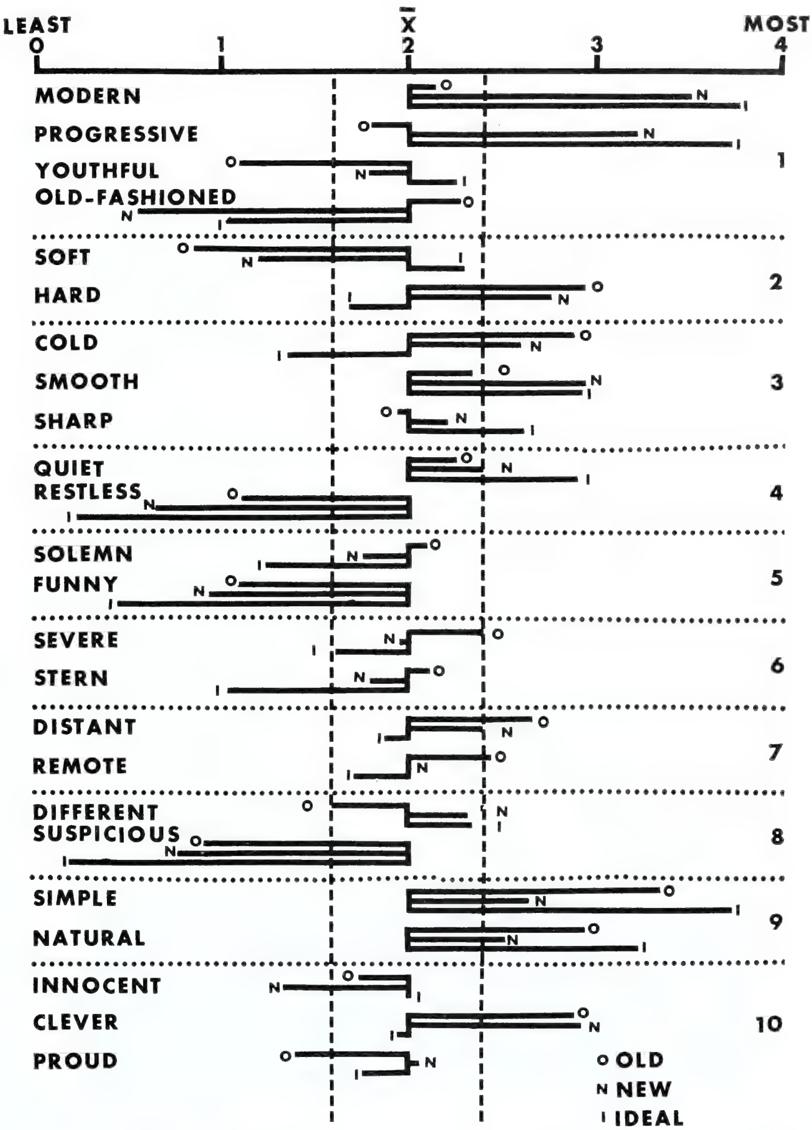


FIG. 33-8 *Qualities for which there is a statistically significant (LC = .95) difference among means for the three types of telephone. In addition, any individual mean above or below the vertical dashed lines is significantly (LC = .95) large or small.*

unsuccessful in satisfying this need. Where the new telephone is far on the other side of the ideal from the old, it is evident that the new design has overshot the mark.

We can consider the distance between the average feeling for the old telephone and the average feeling for an ideal telephone as an indication of the intensity of the need for a design change. Then we might note how close the average score for the new telephone is to the ideal. This can be thought of as an indication of how well the new design satisfies the need for a change.

For all of the qualities of clusters 1 through 8 in Figure 33-8 there is evidence of design improvements. In each case, the new telephone is closer to the ideal than is the old. For example, cluster 1 shows an intense need to improve the modern and progressive qualities of the old telephone. The new design seems to have succeeded in satisfying this need. The new telephone is much closer to the ideal on the three terms "modern," "progressive," and "youthful." However, to some extent the new telephone seems to have overshot the mark in breaking with tradition and is now less "old-fashioned" than the ideal. Cluster 2 exhibits an appreciable need to improve the old telephone's qualities of hardness, or lack of softness. In this case, however, the new telephone is only slightly successful in coming close to the ideal. Cluster 3 indicates a need to improve some of the other structural feelings related to the old telephone. Here the new telephone design produces improved feelings that are almost ideally "smooth," although still somewhat too "cold" and not "sharp" enough. Cluster 4 shows a need for an ideal telephone to have more "quiet" and less "restless" qualities than the old telephone has; the new design achieves about half this objective. In cluster 5 the new telephone is somewhat closer to the ideal in generating feelings that are neither "solemn" nor "funny." Similarly, there are varying degrees of improvement reflected in clusters 6 through 8.

In clusters 9 and 10 the situation tends to be reversed; the old telephone is apparently closer to the ideal than is the new, except for the quality of proudness, for which there is an appreciable overshoot. These design failures seem to revolve around feelings of pretentiousness. That is, the new telephone is felt to be too "proud" and "clever," not "innocent," "simple," or "natural" enough. It is interesting to note that in terms of natural, simple, and innocent feelings the old telephone is rather close to the ideal. In terms of cleverness, both telephones seem far from the ideal. For the quality of proudness, for which there is an overshoot, the old telephone is not proud enough, whereas the new is too proud.

STUDY 3: THE FEELING OF COMFORT

This study was concerned specifically with feelings of physiological comfort as they are influenced by seat design.[5] Because of the contact between the back surfaces of the body and the seat surfaces, the

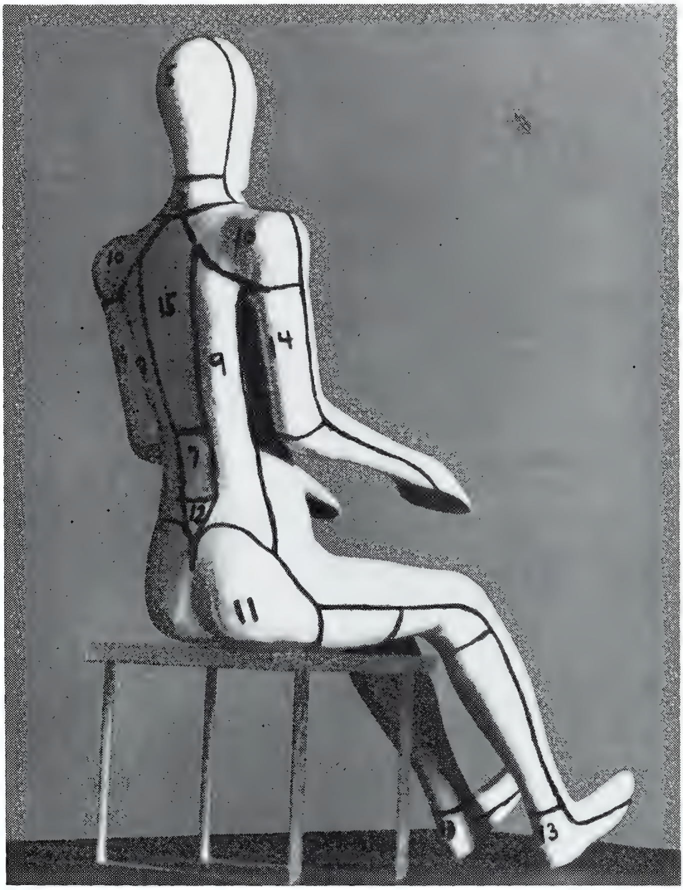


FIG. 33-9 *A manikin used for selecting the most comfortable parts of the body.*

researchers were especially concerned with feelings of comfort as they are patterned across the dorsal portion of the body.

To assess the pattern of comfort, 15 body regions covering the back of the body were delineated as shown in Figure 33-9. These served

[5] This study was done as a thesis by P. S. Allen with the assistance of the author. For complete details on the analyses of these two seats plus four others, see References 1 and 2.

as the set available for choice in the subject's comfort evaluation of the various seat designs. Using these 15 body parts as his guide, a person judged the relative comfort of each part during prolonged sitting in a specific seat.

One problem in judging psychophysiological feelings stems from having to convert the specific body locations of these feelings into words. This was avoided in the present investigation by having the subject examine the 15 areas as they were mapped and numbered on

FIG. 33-10 *Seat C, a Hardman C-124A crew seat.*

the three-dimensional model. The subject did not try to describe his body areas in words. Rather, he was instructed to select from these 15 areas the 3 that he felt were the most comfortable at the moment. With the manikin before him, he recorded the three appropriate code numbers.

After selecting the first 3 areas, he was asked to choose from the remaining 12 the 3 that now were the most comfortable. This procedure continued with a forced choice of 3 from 9 and then 3 from 6, leaving 3 body parts. By this process of multiple forced choice, all 15 body parts were ranked in order of their relative comfort. The first three parts chosen were scored +4, the next three +3, the next three +2, the next three +1, and the remaining three 0. Means were then calculated for each body part.

These data were obtained at fixed ½-hr intervals for a total sitting period of 3½ hr. The scores were averaged over the resulting eight sets of judgments for four male subjects per seat. The four subjects were selected as representing the range of heights of Air Force personnel according to the 1950 survey of flying personnel [12]. One seat, referred to as seat *C*, was a C-124A crew seat. This seat is pictured in Figure

FIG. 33-11 *Seat P, an Aerotherm C-118 pilot seat.*

33-10. A second seat, referred to as seat *P*, was a C-118 pilot seat and is pictured in Figure 33-11.

Results and Conclusions

The results of averaging the eight judgments obtained from the four subjects constrained to each seat for 3½ hr are given in Figures 33-12 and 33-13. The vertical dashed lines indicate the points above or below which a mean is significantly different from the chance mean of +2.00

at the 95 per cent level of confidence. Those means above or below the cutting lines suggest areas of good or bad design, that is, comfort-inducing or comfort-reducing design characteristics.

For seat *C* two significantly good areas and two significantly bad ones are noted. The lower and upper arms feel comfortable. The neck and especially the lower thighs feel uncomfortable. When we examine the picture of the seat, it is immediately evident that the seat does not

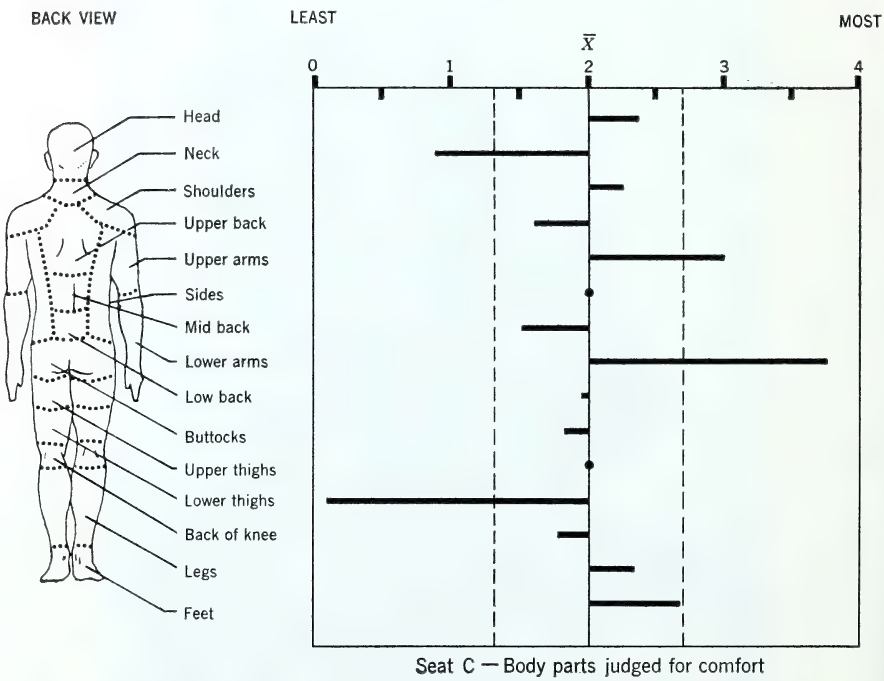


FIG. 33-12 *Pattern of mean judgments of body-part comfort as obtained on seat C. Statistically significant means appear beyond the dashed lines (LC = .95).*

have armrests. The comfort of the arms stems from the absence and not the presence of this feature. Apparently the subjects were able to place their arms in some way acceptable to themselves even though they had no armrests to aid them. The neck area and especially the lower-thigh area have both received some attention from the seat designer, but in both cases the design is inadequate. Obviously the lower-thigh pad is designed incorrectly; it is both too high and too hard. The neck area of the seat seems to slope away from and fails to give adequate support to the neck.

Seat *P*, on the other hand, shows three areas of significant comfort and four areas of significant lack of comfort. The upper arms, the lower arms, and the buttocks are felt to be comfortable. The neck, the mid back the low back, and the feet are not. An examination of the seat shows that there are armrests, apparently of correct design in both height and construction. Even more important, the design has overcome one of the major difficulties of seating in general. That is, there is a comfortable

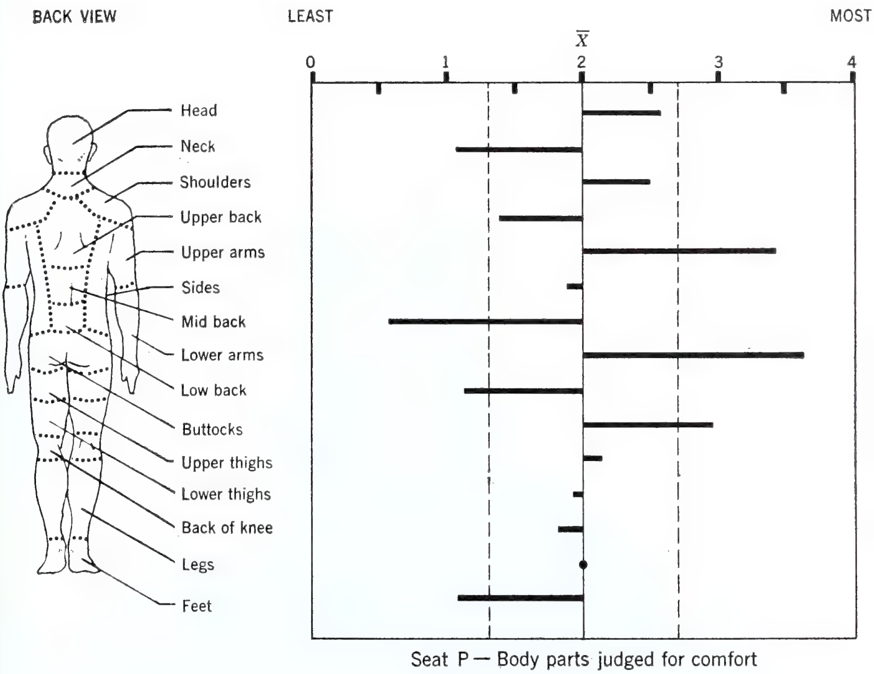


FIG. 33-13 *Pattern of mean judgments of body-part comfort as obtained on seat P. Statistically significant means appear beyond the dashed lines* (*LC* = .95).

place for the buttocks, for at least 3½ hr of sitting. Since there is evidence that the buttocks are a critical focus for seating discomfort [14], this is no small accomplishment. However, there are failures in the design of this seat that are equal in some respects to its successes. The low and mid back are both areas of importance in adequate seating. Seat *P* does not succeed in giving either comfortable low- or mid-back support. The back of the seat apparently slopes away from these areas, and therefore the design fails to meet the need for adequate support. There also is no adequate support for the neck area.

STUDY 4: THE SEATING EXPERIENCE

By using the two seats discussed in study 3 it was possible to study the effects of design on mood or subjective experience.[6] To help clarify this relation, the researchers included a third seat, designed for little more than the rudimentary support of the buttocks and the back. This third seat, which is referred to as seat *X*, is made up of a plywood seat pan fixed at 9° and a plywood back panel fixed at 109°. The seat is pictured in Figure 33-14. It can be considered as a minimum seat, and the subjective experience obtained from sitting in this seat can be used as a base line from which to look at the feelings that other seats produce.

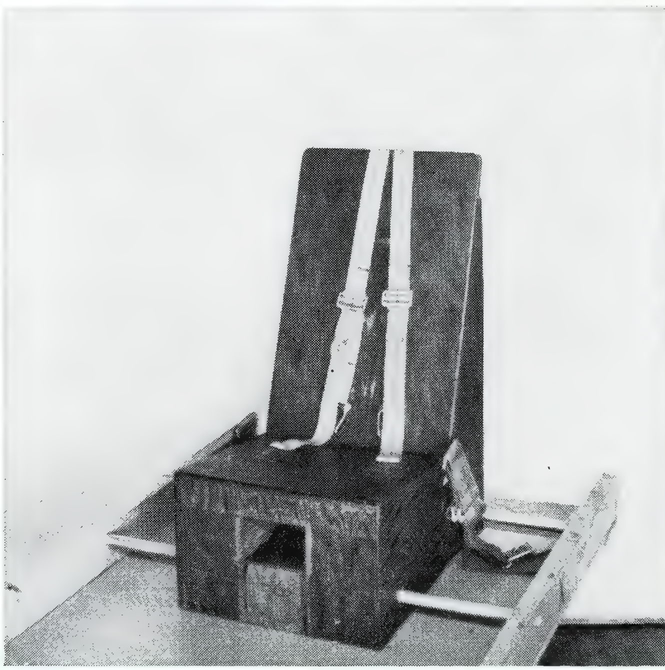


FIG. 33-14 *Seat X, the plywood control seat.*

Four subjects per seat were used, and eight sets of judgments per subject were obtained, one every ½ hr over a period of 3½ hr. The subjects were presented with two previously tested sets of 15 terms each that represented different feeling states.

[6] This study was done as a thesis by P. S. Allen with the assistance of the author. For complete details on the analyses of these two seats plus four others, see References 1 and 2.

One set consisted of the following terms:

Active, alert, calm, comfortable, content, excited, fearless, good, happy, painless, relaxed, restful, right, satisfied, strong.

The other set consisted of the terms:

Passive, tired, disturbed, uncomfortable, discontent, bored, fearful, bad, sad, painful, tense, restless, wrong, dissatisfied, weak.

Together these sets included the positive and negative end points of 15 qualitative continua:

Degree of felt activity, alertness, calmness, comfort, contentment, excitement, fear, value, happiness, pain, relaxation, restfulness, acceptability, satisfaction, strength.

The task involved the multiple forced choice of 3 terms from 15, 3 from 12, 3 from 9, and then 3 from 6, leaving 3. The first three terms chosen were scored +4, the second three +3, and so forth, the last three being scored zero. The subject was presented with one of the sets of 15 terms and asked to select the 3 that most described the way in which the seat made him feel at the moment. This was then repeated with the other sets of terms. The procedure was repeated for the other two seats. The order of presentation for sets of terms and seats was counterbalanced. In this way, comparable information on the subjective feelings that might be related to the three different seat designs was obtained.

Results and Conclusions

The findings that relate to the three seats are shown in Figures 33-15 and 33-16 (seat *X*), Figures 33-17 and 33-18 (seat *C*), and Figures 33-19 and 33-20 (seat *P*). The vertical dashed lines indicate the points above or below which a mean is significantly different from the chance mean of +2.00 at the 95 per cent level of confidence.

An inspection of Figures 33-15 and 33-16 suggests that the basic plywood-seating experience includes strong feelings of discomfort, restlessness, dissatisfaction, and pain. These are the negative feelings to be reduced by improved seat design. On the other hand, this basic plywood-seating experience also includes positive feelings of social propriety (rightness and goodness), strength, and fearlessness. These may be worthy feelings that should not be lost by improper design modifications.

It would appear, as evidenced in Figures 33-17 and 33-18, that the design of seat *C* fails in all respects to improve the seating experience. In addition, there are errors of design that apparently increased feelings

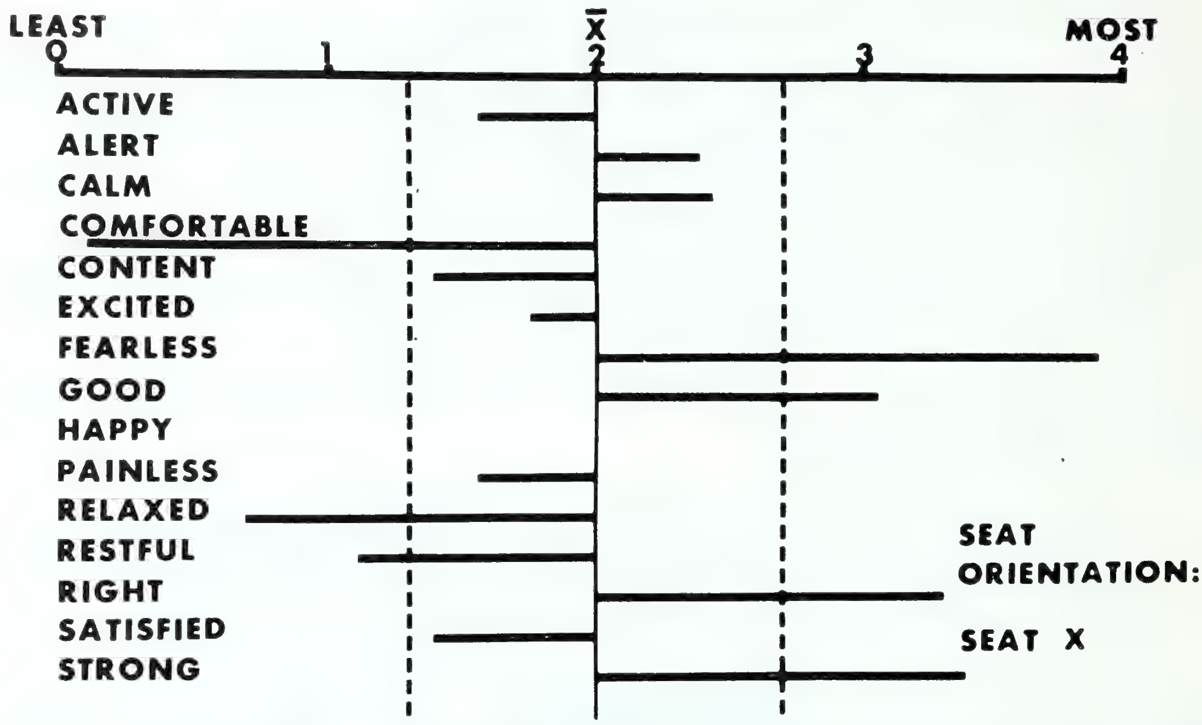


FIG. 33-15 *Pattern of mean judgments of feeling obtained on seat X. Positive terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

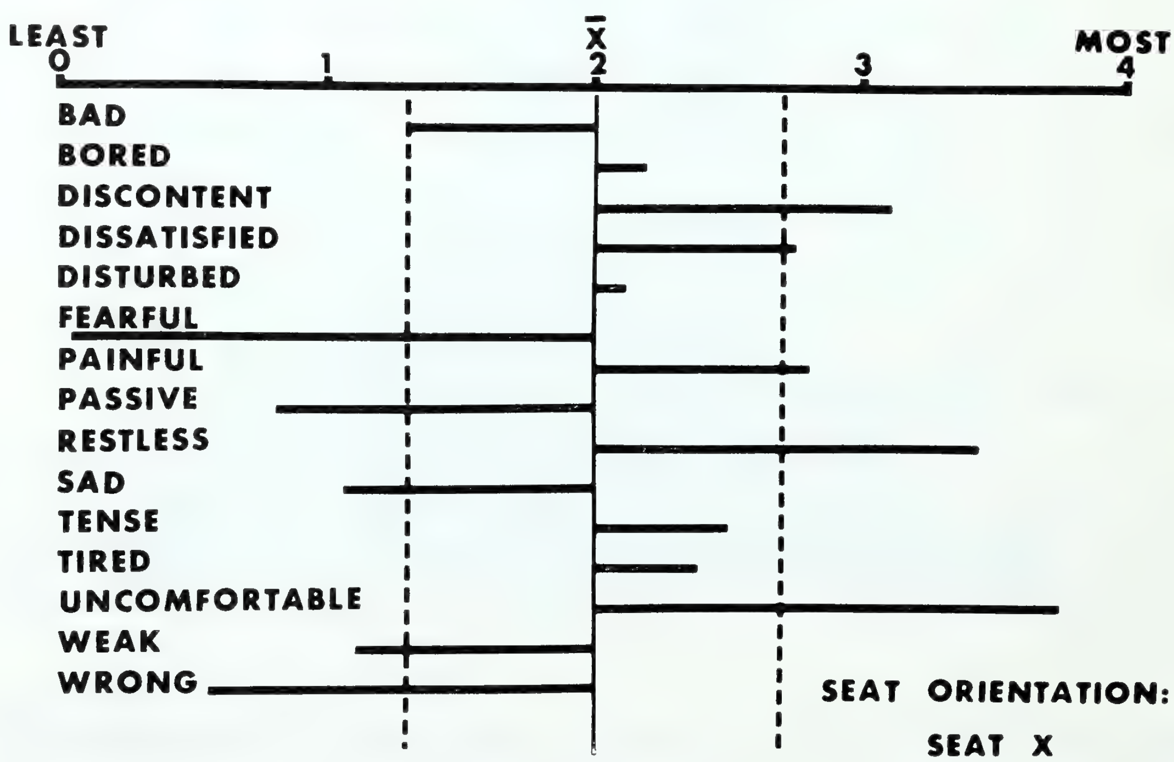


FIG. 33-16 *Pattern of mean judgments of feeling obtained on seat X. Negative terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

of discontent, disturbance, and fatigue. The seat also seems to neutralize the positive feelings of strength, fearlessness, and propriety.

On the other hand, the design of seat *P*, as indicated in Figures 33-19 and 33-20, succeeds in two important areas. It reduces feelings of discomfort and pain. At the same time, however, it fails in that it results

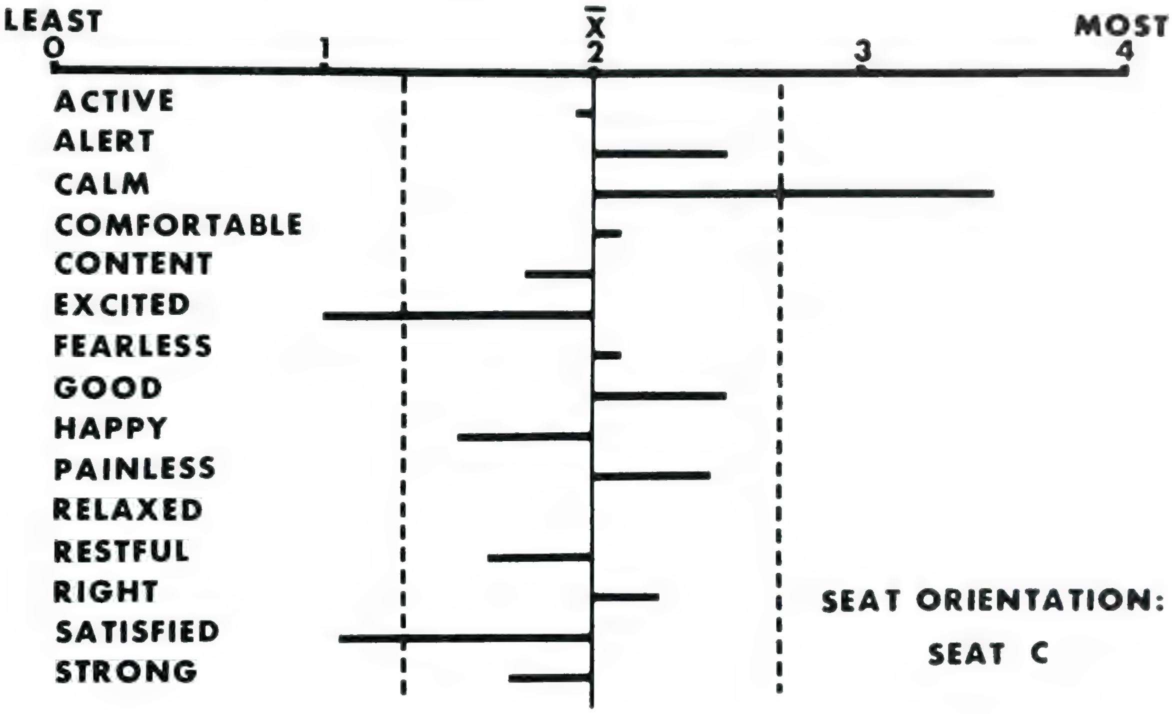


FIG. 33-17 *Pattern of mean judgments of feeling obtained on seat C. Positive terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

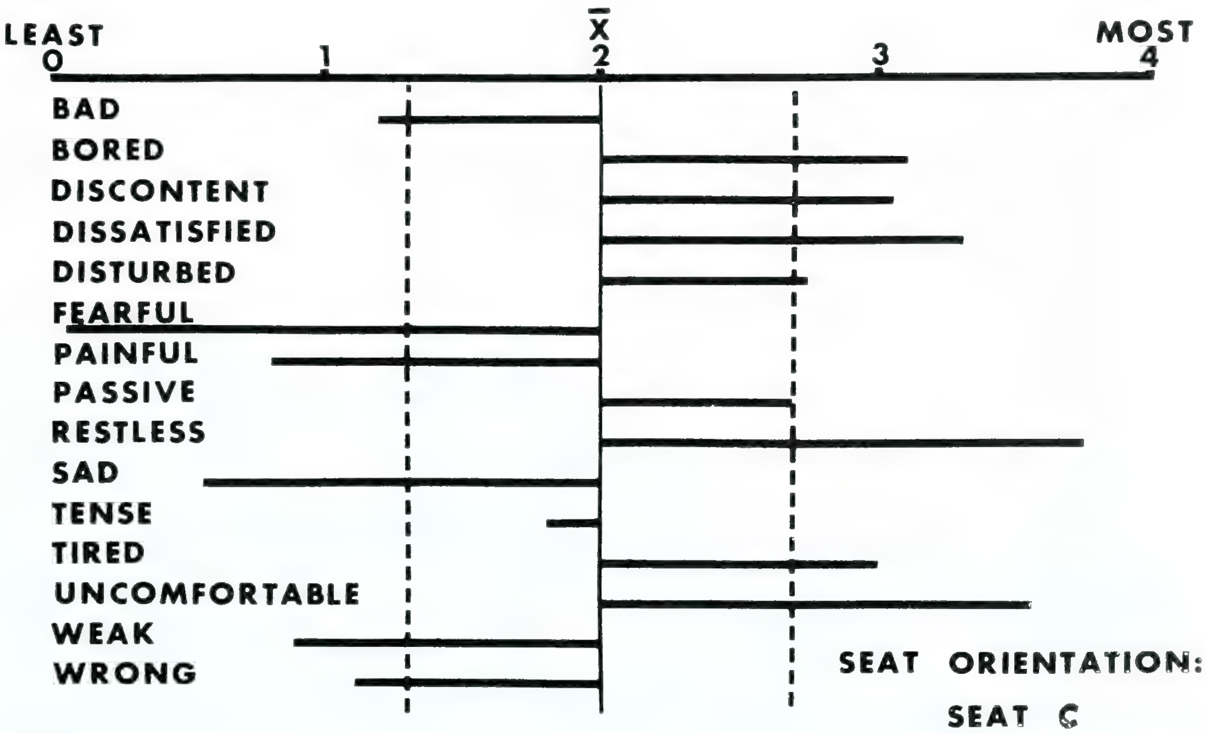


FIG. 33-18 *Pattern of mean judgments of feeling obtained on seat C. Negative terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

in higher feelings of restlessness. The design apparently does not neutralize either the feelings of fearlessness or some of the feeling of social propriety. The feeling of strength, however, does tend to disappear.

In general, on the basis of this form of data, we can suggest that an ideal seat would be one in which the negative feelings evidenced on

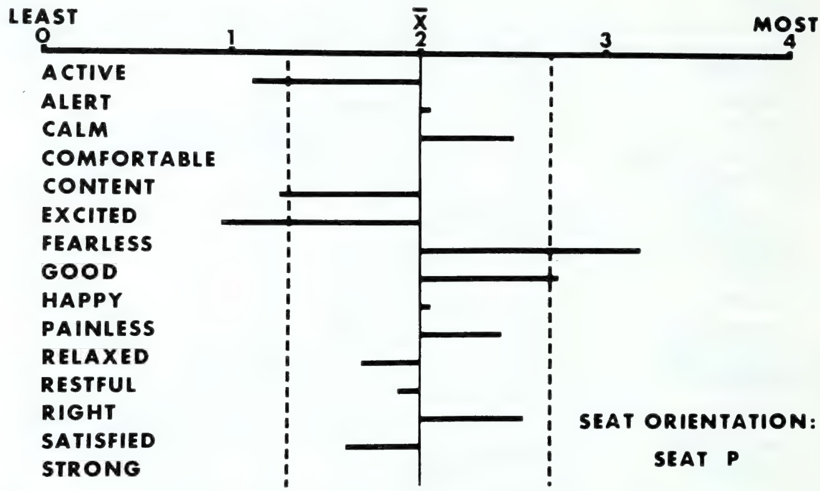


FIG. 33-19 *Pattern of mean judgments of feeling obtained on seat P. Positive terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

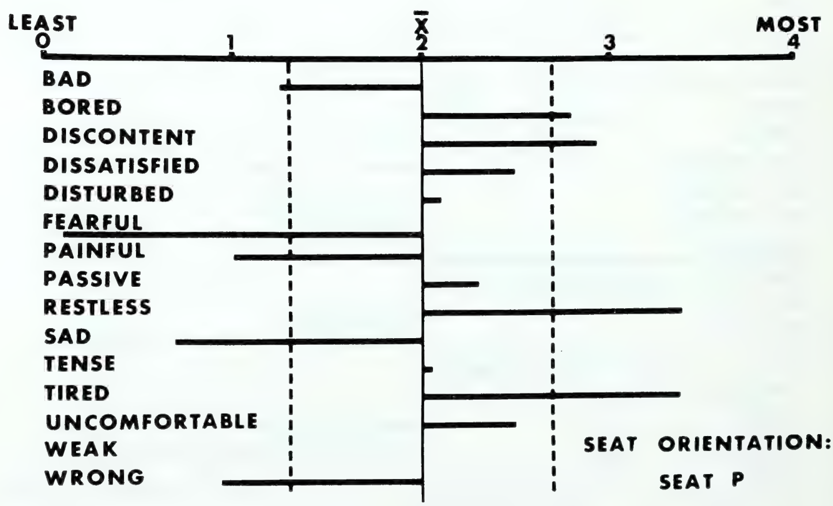


FIG. 33-20 *Pattern of mean judgments of feeling obtained on seat P. Negative terms. Statistically significant means appear beyond the dashed lines (LC = .95).*

the basic plywood seat *X* all disappear, leaving the positive feelings that we wish to keep. We might even add positive feelings which are not natural to the seating experience but which are desirable. In any case, it would seem that a good seating experience would, at least, not induce discomfort, dissatisfaction, restlessness, or pain and would not reduce one's sense of social propriety.

GENERAL APPLICATIONS

In general, there are two types of analyses that can be used in studying technological design and product effects. Each supplies a different kind of information, and both have their places at one or another time in the development or merchandising of a product.

Intraproduct Analysis

This type of investigation explores the high points and low points in the personality of a product, or in the user's reaction to a product, or in feelings resulting from the use of a product. It should be noted that the analysis takes place only within the framework of one product. It calls for an analysis of the feelings toward that specific product by comparison with chance feelings. That is, a feeling is significantly high or low only so long as it is greater or less than could be expected on the basis of chance. It is not necessarily higher or lower than the feeling for some other product.

For this analysis, the sets of response terms are formed along established lines, with as much caution and preliminary thinking as possible. When the appropriate sets of terms have been designed, or when already developed and pretested sets [4] are being used, appropriate questions are prepared. These can take many forms, depending upon what we want to know. For example:

1. How do you feel about _____?
2. How does _____ make you feel?
3. How would you describe _____?
4. What does _____ remind you of?
5. How do you feel now, just having used _____?

Occasionally we may be interested in the pattern of feelings about the users of the product as well as about the product itself. In this case, we might ask:

6. How would you describe a person who uses _____?
7. How do you think you would feel if you were to use _____?

Interproduct Analysis

This form of analysis calls for comparing one product with another. All other conditions are the same as those in an intraproduct analysis, except that the data are collected for more than one product. Critical

areas are then noted where there is a statistically significant difference between what is felt for one product and what is felt for another. It is possible to introduce as many products into the analysis as is desired. It is also possible to introduce an "ideal" product, as was done in study 2, or a "basic" product, as was done in study 4. In any case, one precaution must be taken. The administration, the sets, and the instructions must be identical in all respects. Only the products should be different. Otherwise the comparison cannot be clearly interpreted.

An interproduct analysis provides all the data required for an intraproduct analysis. Therefore, often the two analyses are done together, giving a combined result. The intraproduct analysis gives us absolute data, that is, highs and lows independent of anything else. The interproduct analysis adds relative data, highs and lows by comparison with some other thing.

When the results warrant the effort, one additional approach is to use more than one product and more than one group of subjects. For example, we might select four groups, (1) those who use or like *X*, (2) those who use or like *Y*, (3) those who refuse to use or who dislike *X*, and (4) those who refuse to use or who dislike *Y*. Then we should obtain data from all four groups in respect to both product *X* and product *Y*. A comparison of results from such an investigation can further pinpoint areas of good and bad design in competitive products.[7]

BIBLIOGRAPHY

1. Allen, P. S.: Forced-Choice Rankings as a Method for Evaluating Psycho-Physiological Feelings, master's thesis, Tufts University, Medford, Mass., 1958.

2. ———, E. M. Bennett, and D. K. Kemler: Forced-choice Ranking as a Method for Evaluating Psychophysiological Feelings, *USAF WADC Tech. Rept.* 58-310, 1958.

3. Bennett, E. M.: Empirical Aspects of Polydiagnostic Research, *J. Psychol.*, vol. 44, pp. 251–269, 1957.

4. ———: "Personality Assessment and Diagnosis," The Ronald Press Company, New York, 1961.

5. ———: The Polydiagnostic Method: A Technique for Multivariate Social and Clinical Research, *J. Psychol.*, vol. 42, pp. 207–215, 1956.

6. ——— and L. R. Cohen: Men and Women: Personality Patterns and Contrasts, *Genet. Psychol. Mon.*, vol. 59, pp. 101–155, 1959.

7. ———, ———, and D. K. Kemler: Preliminary Standardization of the Polydiagnostic Index, *J. Psychol.*, vol. 43, pp. 307–324, 1957.

[7] See Reference 8 for a detailed example of this complex analysis as applied in a political study.

8. ——— and H. Goodwin: Emotional Aspects of Political Behavior: The Woman Voter, *Genet. Psychol. Mon.*, vol. 58, pp. 3–53, 1958.

9. ———, D. K. Kemler, and B. T. Levin: Emotional Associations with Air and Rail Transportation, *J. Psychol.*, vol. 43, pp. 65–75, 1957.

10. Forrest, J., and E. M. Bennett: Esthetics Research in Human Design, paper presented at American Psychological Association Convention, New York, September, 1957.

11. Guilford, J. P.: "Psychometric Methods," McGraw-Hill Book Company, Inc., New York, 1954.

12. Hertzberg, H. T. E., and G. S. Daniels: Anthropometry of Flying Personnel—1950, *USAF WADC Tech. Rept.* 52-321, 1954.

13. Johannsen, D. E.: The Principle of Psychophysics with Laboratory Exercises, Author, Saratoga Springs, N.Y., 1941.

14. Slechta, R. F., and E. A. Wade: Comparative Evaluation of Aircraft Seat Cushions, *USAF WADC Tech. Rept.* 58-311, 1958.

15. Stevens, S. S.: "Handbook of Experimental Psychology," John Wiley & Sons, Inc., New York, 1951.

Chapter 34

ANALYTIC TECHNIQUES IN THE DEVELOPMENT OF CONTROLS AND DISPLAYS FOR ORBITAL FLIGHT

*Charles Owen Hopkins**

THE FOLLOWING ANALYSIS ILLUSTRATES HOW DESIGNS FOR DISPLAYS AND controls to be used during orbital flight are developed. The development steps include (1) mission and systems analyses to determine the necessary system functions, (2) feasibility studies to evaluate alternative methods of performing system functions, (3) mission-profile–system-function synthesis to identify system functions to be performed by the human operator, (4) study of information required by the human being in order to perform the functions assigned to him, and (5) design of displays and controls.

The problem of assigning system functions to human operators is not easily solved. The guiding principle is the utilization of the man in the performance of those functions which he handles better than the automatic devices which are available. What functions the man performs better than a machine cannot be specified completely in quantitative terms; there is no set of equations that represents all the relations between human inputs and outputs. Moreover, the choice between a human operator and an automatic device often depends upon peculiar system requirements rather than upon the relative efficiency of performance under ideal conditions. Therefore, the system functions that should be performed by a human operator can be determined only within the context of a specific system intended to accomplish a particular mission. To determine what the human being should do, it is necessary to know what the system must do. What the system must do is determined by the mission requirements.

* Hughes Aircraft Company, Culver City, Calif.

This research, performed in 1958–1959, was supported in part by the U.S. Air Force under Contract AF33(616)-6033, monitored by Aerospace Medical Laboratory, Directorate of Laboratories, Wright Air Development Division, Wright-Patterson Air Force Base, Ohio.

TYPICAL MISSION ANALYSIS

Mission Purpose

The purpose is that of transporting personnel, equipment, and supplies between the earth and an earth-orbiting space station. A ferry vehicle is injected into a planned orbit from a ground launch. It executes an orbit transfer to rendezvous with the satellite already in orbit. After transferring stores and/or personnel, the ferry departs from orbit, reenters the earth's atmosphere, and glides to a landing at a designated base.

Assumptions and Constraints

The following assumptions and constraints are applicable to the mission:

1. The destination satellite is in a polar circular orbit at an altitude of 500 nautical miles.
2. From the ground launch the ferry vehicle is injected into a circular orbit at an altitude of approximately 300 nautical miles. This initial orbit is approximately coplanar with the destination-satellite orbit.
3. A minor change in the plane of the initial orbit is considered a normal maneuver that occurs as a part of the planned mission.
4. All orbital maneuvers, including plane change, orbit transfer, and departure from orbit, are accomplished by means of programs requiring minimum energy expenditure.

Major Mission Requirements

These requirements are derived from the mission purpose and the assumptions and constraints. They may be considered as subordinate goals that must be achieved for successful completion of the mission. In their order of occurrence, these requirements are:

1. Initiation of powered flight
2. Attainment of the planned orbit
3. Rendezvous with the destination satellite
4. Departure from orbit
5. Attainment of aerodynamic flight
6. Arrival in the vicinity of landing base
7. Controlled landing at the predesignated base

Since the accomplishment of each major requirement marks the end of a sequence of idiosyncratic events, the mission may be divided

into the following phases: prelaunch, boost, rendezvous, deorbit, reentry, terminal navigation, landing.

A key event is selected to signify achievement of each major mission requirement. For example, achievement of the first requirement, initiation of powered flight, may be signified by the delivery of the firing signal to the first-stage rocket motor. The conditions necessary for the event to occur are then specified. For the initiation of powered flight, for instance, the necessary conditions are (1) the time to go to predetermined optimum launch time equals zero and (2) the pilot and ground crew agree that the system is ready for launching.

The mission analysis is summarized in Table 34-1. Since this

TABLE 34-1 *Summary of Mission Analysis*

Major mission requirement	Mission phase terminated by accomplishment of requirement	Key event signifying accomplishment of mission requirement	Necessary conditions for occurrence of event signifying accomplishment of mission requirement
Initiation of powered flight	Prelaunch	Delivery of firing signal to first-stage rocket motor	Zero time to go to predetermined launch time; pilot and ground crew agreement on system readiness for launch
Attainment of planned orbit	Boost	Burnout of injection-stage rocket	Velocity, altitude, and velocity vector within specified limits
Rendezvous with destination satellite	Rendezvous	Arrival so as to remain within arbitrary range of satellite	Correspondence, within close tolerances, of orbital elements of the two satellites
Departure from orbit	Deorbit	Burnout of retro rocket	Burnout at correct time; speed, altitude, and velocity vector within specified limits
Attainment of aerodynamic flight	Reentry	Attainment of specified "safe" velocity	Reentry into atmosphere at small, "safe" angle; glide program to reduce orbital speed and avoid overheating
Arrival in vicinity of landing base	Terminal navigation	Arrival at entry notch within acceptable velocity and altitude limits	Attitude-control program as a function of velocity and geographical position during atmospheric glide
Controlled landing at designated base	Landing	Termination of landing roll or slide in safe condition	Attitude-control program as a function of velocity and geographical position during landing approach

example is limited to the requirements for orbital-flight displays and controls, the subsequent discussion is confined to the rendezvous phase.

The key event signaling the accomplishment of rendezvous may be either safe mechanical attachment to the destination satellite or arrival so as to remain in the vicinity of the satellite. For the present mission it is assumed that the requirement is that of "parking" rather than mechanical attachment. The conditions necessary for the occurrence of this event are (1) approach to within a certain range and (2) matching of the orbital elements of the two satellites within specified small tolerances. If the periods of the two satellites are matched and there are only minor differences in some of the orbital elements, the ferry will remain in the vicinity of the target, although its position relative to the target and its range from the target will vary as the two satellites orbit the earth.

SYSTEMS ANALYSES

System Requirements

The next step is to specify the system requirements that must be met in order to achieve the conditions necessary for the occurrence of the key event. These requirements are valid for all vehicles of the class under consideration, regardless of differences in detailed design characteristics.

Stabilize Attitude. The first system requirement following cutoff of the injection-stage rocket motor is that of stabilizing vehicle attitude, which involves either the securing of equilibrium from cutoff conditions consisting of angular and/or rate errors or the maintaining of the equilibrium associated with efficient, continuous control. Uneven cutoff conditions may result in tendencies toward attitude errors or rates or certain combinations of both. Normally, attitude will be controlled continuously, either automatically or manually, during burning and cutoff of the injection-stage rocket. If there is efficient, continuous control, attitude rates of any appreciable magnitude may never occur.

Initiate Pitch Rate. After the vehicle attitude has been stabilized, the next requirement is for initiating a pitch rate to keep the longitudinal axis of the vehicle perpendicular to local vertical. For a condition of zero yaw error, this may be accomplished by initiating a pitch rate equal to the angular rate of local vertical. For circular orbits, the angular rate of local vertical is constant at a value determined by the radius of the orbit. Therefore, for a circular orbit at any specified altitude, a pitch rate that produces one complete rotation of the vehicle around its pitch axis (when the pitch axis is perpendicular to the plane of the orbit)

during each orbital period will keep the "bottom" of the vehicle parallel to the surface of the earth.

Determine Orbital Elements. If the stable platform and digital computer have been functioning properly, the orbital elements and related parameters will have been determined immediately upon injection into orbit. If either of these components malfunctions, the system requirement following establishment of pitch rate is determination of the orbital elements. Even if this requirement is met automatically at injection-stage cutoff or manually or semiautomatically after initiation of an appropriate pitch rate, it is a prerequisite for successful execution of any subsequent orbital maneuvers, i.e., orbit-plane rotation, orbit transfer, and deorbit.

Rotate Orbit Plane. The next system requirement is that of rotating the orbital plane. A minor plane change for making the orbit coplanar with the destination satellite is considered to be a normal maneuver required as a part of each mission. In order to execute the maneuver with minimum expenditure of energy, corrective thrust will be applied at one of the two points where the vehicle orbit intersects the plane of the destination-satellite orbit.

Transfer to Destination-satellite Orbit. After the two orbits are approximately coplanar, it is necessary to transfer from the inner orbit to the outer orbit in order to intercept the destination satellite. Minimum energy transfer is accomplished by means of a Hohmann transfer orbit, an elliptical orbit cotangential with the ferry and destination-satellite orbits. Two thrust applications are required. The first is applied when the central angular relation between the two vehicles is such that the ferry will arrive at the apogee of its transfer orbit at the same time as the destination satellite arrives at that point in its circular orbit. The second thrust is applied at the apogee of the transfer ellipse to achieve circular velocity at that altitude.

System Functions

Certain functions must be performed by the system if it is to meet the five system requirements. For example, in order to stabilize attitude, the system must (1) sense vehicle-attitude conditions, (2) determine corrective attitude jet thrust, and (3) apply corrective thrust. The system functions necessary for each of the requirements during the rendezvous phase are summarized in Table 34-2.

Feasibility Studies

The system functions may be performed in a number of different ways, which vary in degree of accuracy, reliability, and safety and in weight and volume of equipment required. Functional-error analyses

TABLE 34-2 *System Requirements and System Functions during Rendezvous Phase*

System requirements	System functions
Stabilize attitude	Sense vehicle-attitude conditions
	Determine corrective attitude jet thrust
	Apply corrective thrust
Initiate pitch rate	Determine correct pitch rate
	Determine required thrust
	Apply thrust to initiate pitch rate
Determine orbital elements	Determine orbital elements or related parameters
	1. Size and shape of orbit
	2. Position of vehicle in orbit
	3. Orientation of orbit in space
Rotate orbit plane	Determine time of arrival at node
	Determine amount and direction of required thrust
	Apply thrust to rotate plane of orbit
	Repeat attitude stabilization if necessary
Transfer to destination-satellite orbit	Determine critical central angle between satellites
	Determine time of occurrence of critical central angle
	Determine required thrust magnitude
	Apply thrust to initiate transfer ellipse
	Repeat attitude stabilization if necessary
	Determine arrival time at apogee of transfer ellipse
	Apply thrust at apogee to achieve circular velocity
	Repeat attitude stabilization if necessary

and tradeoff studies serve to determine the better methods. Further choice is based upon a preference for components and subsystems that may be used in the performance of system functions during several mission phases.

ALLOCATION OF SYSTEM FUNCTIONS

A provisional assemblage of components, including the human elements, that is considered capable of performing the required system functions is used as a basis for allocating functions to the human operator. The major system components involved in the performance of rendezvous-phase functions are presented in Table 34-3. Because it is

TABLE 34-3 *Major System Components for Performance of Rendezvous-phase Functions*

Primary components	Auxiliary components	Miscellaneous components
Stable platform	Radar altimeter	Clock
Digital computer	Horizon scanner	Interval timer
Automatic attitude-stabilization system (AASS)	Star tracker	Orientable periscope
Human components	Yawscope	Tracking radar
		Graphs and tables

descriptive of a self-sufficient system the table lists no provision for the acquisition of information from either earth-based tracking and computing centers or other satellites.

Mission-Profile–System-Function Synthesis

For the first step in constructing a diagram of a mission-profile–system-function synthesis, each of the three major electromechanical components of the system is considered to be in either of two states: operative or inoperative. Therefore, as shown in Figure 34-1, the diagram

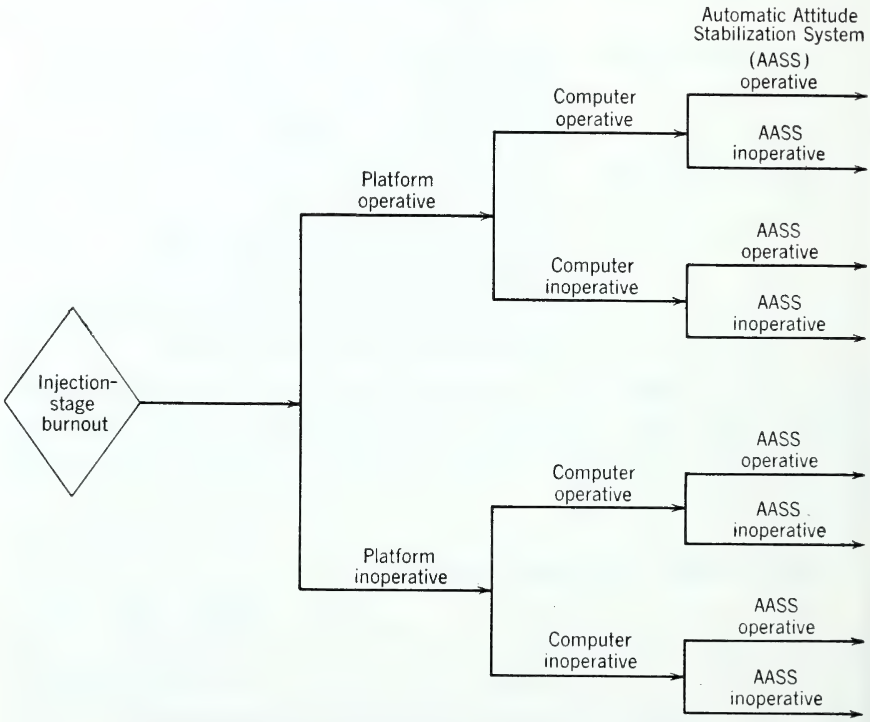


FIG. 34-1 *Potential event sequences at injection-stage burnout.*

potentially is composed of eight alternative event sequences, beginning at injection-stage burnout, that are each determined by a particular combination of operative and inoperative system components.

Three kinds of events depicted in this diagram are then coded with geometrical symbols. An action carried out automatically is shown by a diamond. An action initiated or carried out by the human operator is shown by a rectangle. Information that is received by the human operator is shown by a circle. A new diagram is constructed to illustrate

how the necessary system functions are accomplished with various combinations of system components, including the human operator and the auxiliary components. Any one of the eight basic event sequences may either branch into additional event sequences or terminate, depending upon the occurrence of unfavorable events and the possibility or impossibility of coping with the unfavorable event by means of alternative components, alternative modes of operation, or alternative mission goals. An example of a mission-profile–system-function synthesis diagram is shown in Figure 34-2. This diagram illustrates the performance of the system functions of (1) stabilizing attitude and (2) initiating pitch rate. The level of description of events is rather general, since at this point the specific configurations of displays and controls have not been determined.

Summary of Human Functions during Rendezvous Phase

The system functions that are to be performed by the human operator may be identified from the diagrams and classified into the following general categories:

1. Selection of one of a number of alternate sensing devices to be used for acquiring information
2. Manual operation of certain sensing devices
3. Transmission of manually acquired information to the system
4. Limited processing of data by means of charts, graphs, tables, and, in some cases, a special-purpose hand-operated computer
5. Manual control of vehicle output in certain modes of operation
6. Evaluation of the state of the system in the light of planned mission requirements and the occurrence of unfavorable events
7. Designation and, in some cases, execution of alternative basic system programs as a result of the evaluation of the state of the system

DISPLAYS AND CONTROLS

Each human function included in the above seven categories requires certain kinds of information for its proper execution. The principal categories of required information are (1) attitude information, (2) orbital-status information with reference to (*a*) the earth and (*b*) the destination satellite, and (3) thrust-control information. For each of these three types of information it is necessary to provide (1) requirement or command information, (2) capability information, and (3) component- and subsystem-status information. An example of specific

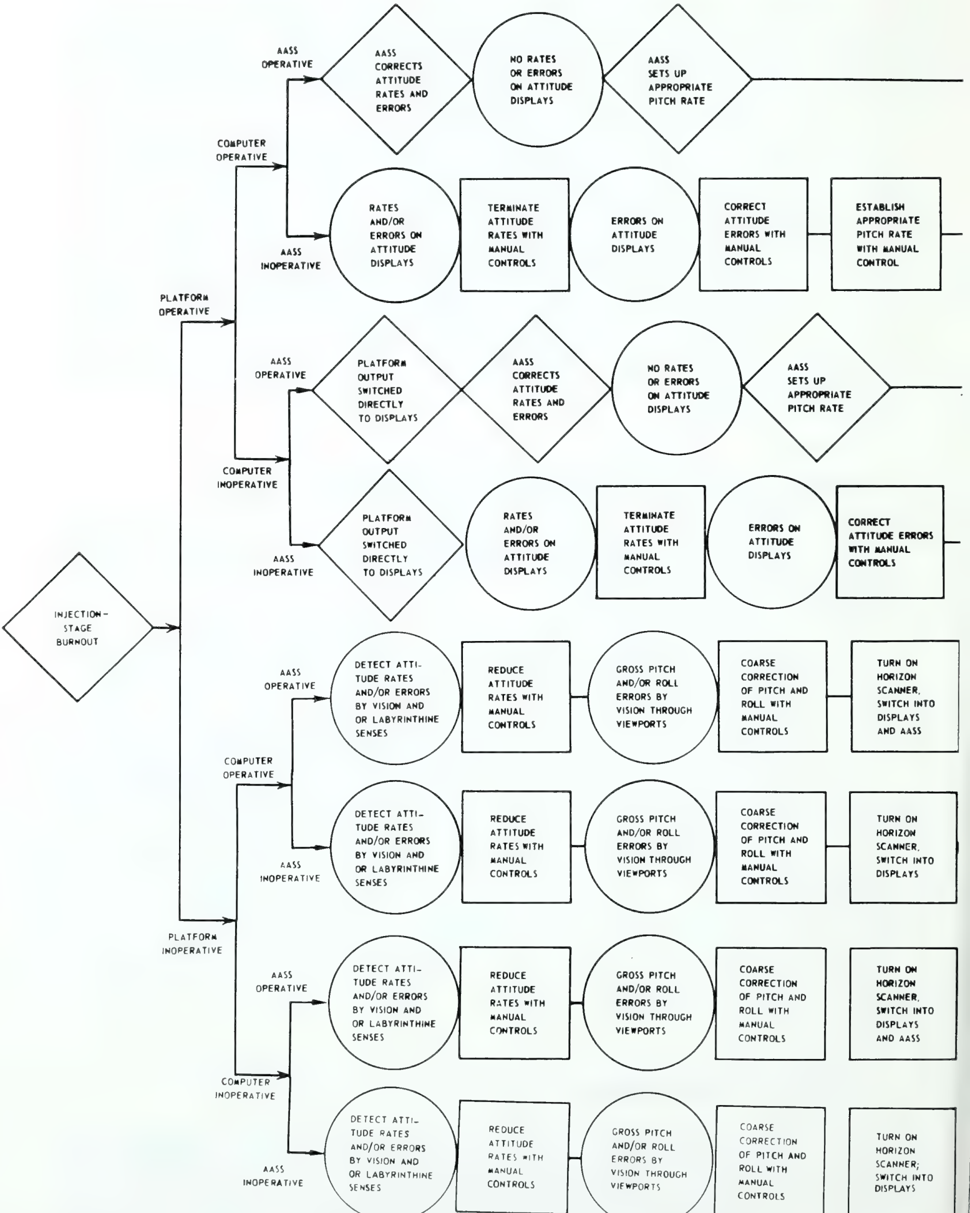
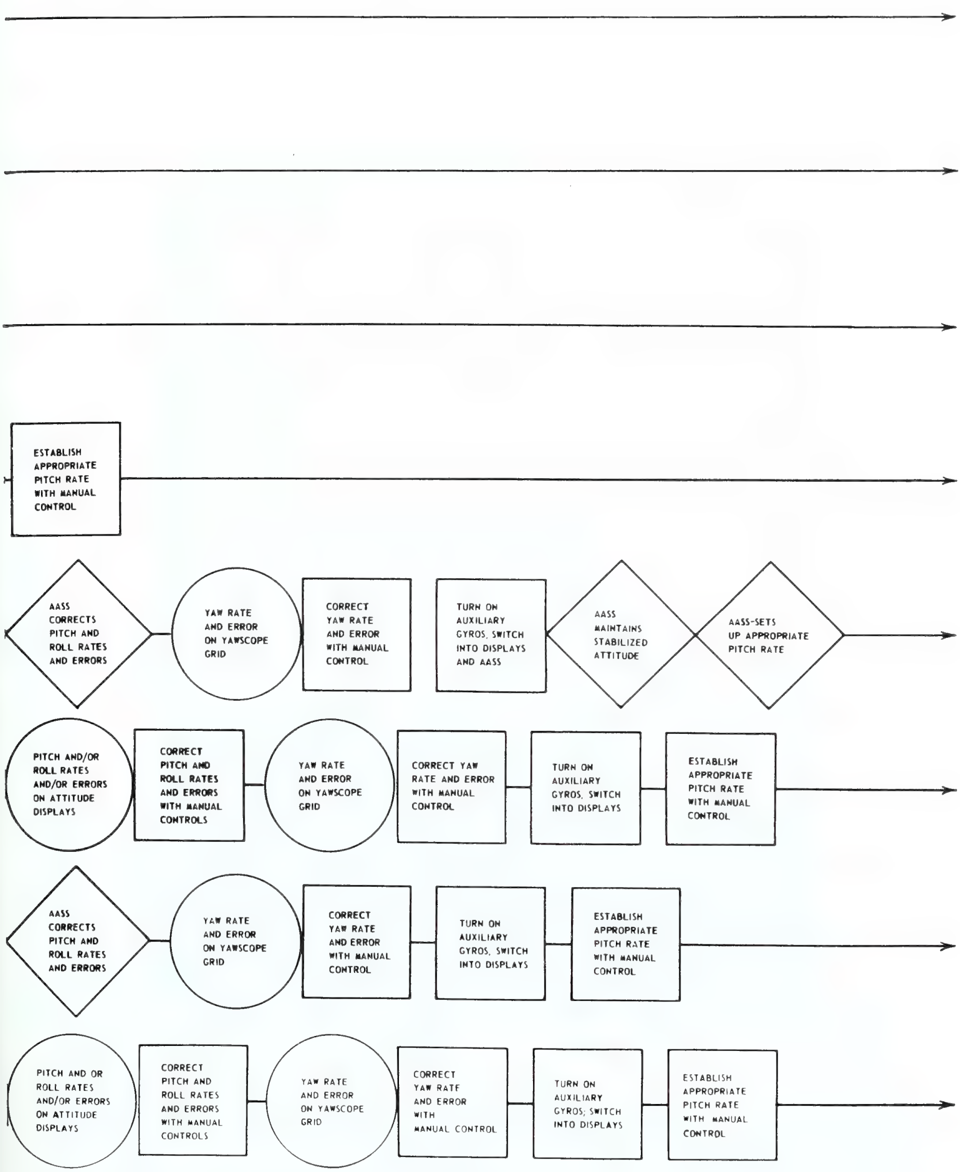


FIG. 34-2 *Mission-profile–system-function synthesis diagram*



of stabilization of attitude and initiation of pitch rate.

information requirements is shown in Table 34-4. Similar lists are prepared for all categories of required information and are used as the basis for preliminary design of displays and controls.

TABLE 34-4 *Summary of Orbital-status Information and Controls with Reference to Target Satellite*

Primary information	Requirement and command information	Capability information	Component- and subsystem-status information	Associated controls
Range to satellite 1. Radar 2. Optical Radar azimuth and elevation tracking angles	....	1. Predicted minimum range 2. Collision warning	Radar status	1. Radar power 2. Stadiametric ranging
Pictorial analog of orbital relations between the two satellites	....	1. Predicted positions relative to satellite for next period 2. Predicted range to satellite as a function of time for duration of rendezvous		
Horizontal and vertical optical look angles to satellite Visual image of satellite	Position of satellite relative to ferry	....	....	1. Unstow periscope or TV camera 2. Orient periscope or TV camera

Some display and control designs used in two full-scale mockups of cockpits for two alternative display and control systems are shown in Figures 34-3 and 34-4. Two of these displays will be described in detail in order to indicate the final result of the previously described analyses.

The two primary display areas in each of the two cockpits are time-shared among different displays. One of these, the 10-in.-square display surface, located in the center of the inclined front panel in both cockpits, is time-shared for optical projection-type displays, cathode-ray-tube displays, and a periscopic viewing screen. The two displays to be described are the transfer display and the rendezvous display. These displays differ only in scale factor and the times when they are used.

Transfer Display

The transfer display, shown in Figure 34-5, is a cathode-ray-tube presentation that provides information about the position and predicted

path of the ferry vehicle relative to the destination satellite. The display is used while the ferry is in the transfer orbit.

The reference system used for specifying the position of the ferry is a moving cartesian-coordinate system with its origin at the center of the destination satellite. The y axis is directed along the satellite's radius

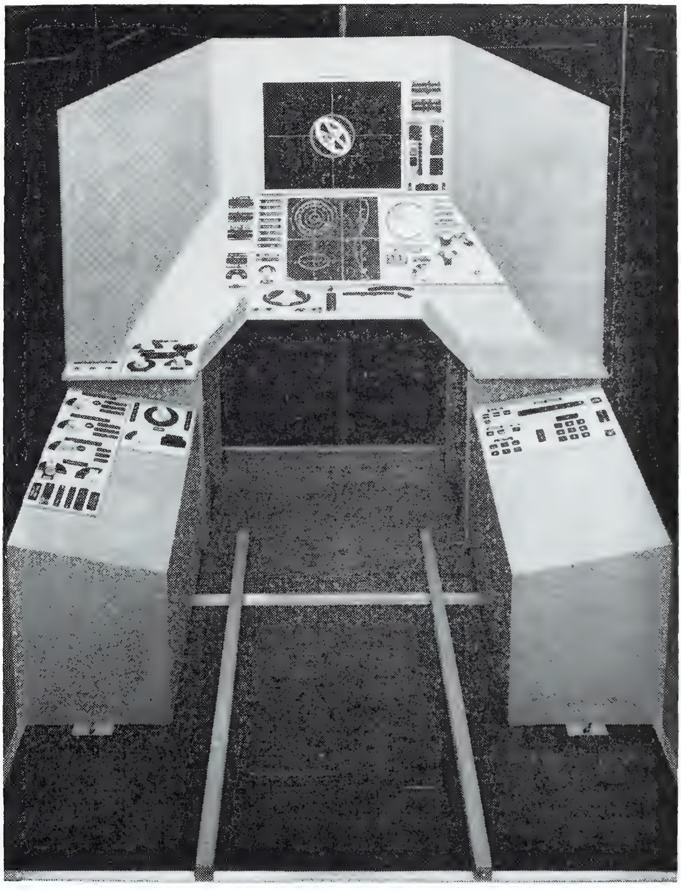


FIG. 34-3 *Orbital-vehicle cockpit mockup (Mark I).*

vector. The x axis is normal to the radius vector in the plane of the orbit; the z axis is normal to the plane of the orbit. The position of the destination satellite in its orbit is specified in degrees of central angle (η) measured from the ascending mode in the direction of satellite travel.

A rectangular area 5 in. high and 6 in. wide, located in the lower left part of the display, is devoted to an X-Y presentation. The predicted path of the ferry vehicle relative to the destination satellite, projected

onto the plane of the destination-satellite orbit, is shown in the X-Y presentation. The vertical and horizontal lines represent the x- and y-axes of the coordinate system. Index marks spaced ½ in. apart along these lines represent distance increments of 10,000 ft along the axes. The small circle at the intersection represents the destination satellite.

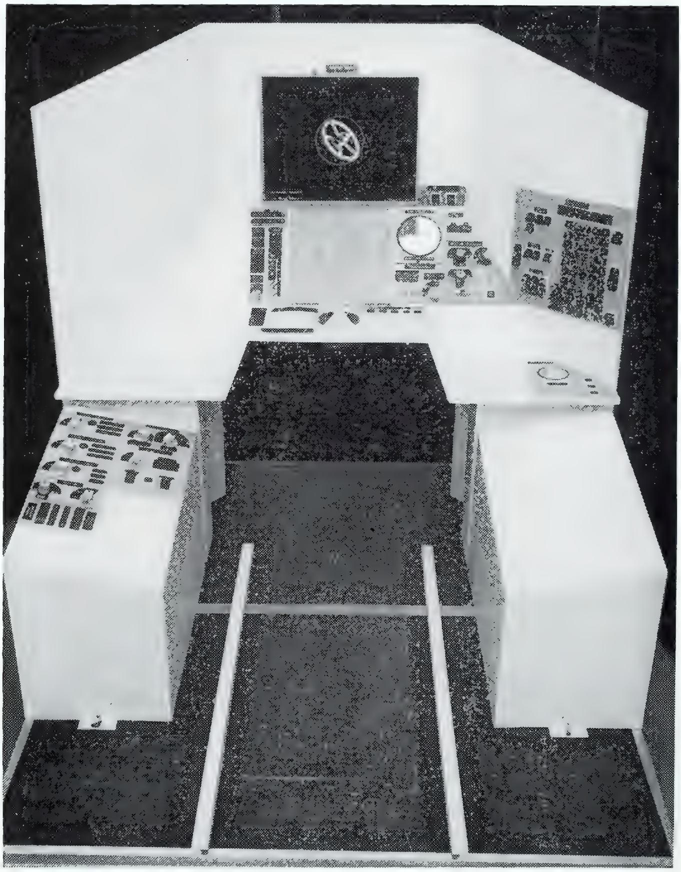


FIG. 34-4 *Orbital-vehicle cockpit mockup (Mark II).*

The predicted path of the ferry relative to the destination satellite, as computed from the orbital elements of both satellites, is shown by a curved cathode-ray trace against the fixed reference axes. The apogee of the transfer orbit, which is also the predicted point of thrust application for attainment of circular orbital velocity at the altitude of the target satellite, is shown by a short bar intersecting the predicted path at a right angle. (In Figure 34-5 this bar coincides with the vertical axis.)

A Z-η presentation occupies a rectangular area 4 in. wide and 10 in. high at the right side of the display. The vertical line represents the plane of the destination-satellite orbit; the horizontal line represents the earth's equatorial plane. The intersection of these lines represents the ascending mode of the destination-satellite orbit. The scale on the vertical axis represents degrees of central angle measured from the ascending mode in the direction of satellite travel. The index marks on the horizontal axis represent distance increments of 10,000 ft from the plane of the orbit in a direction normal to the orbit (z scale).

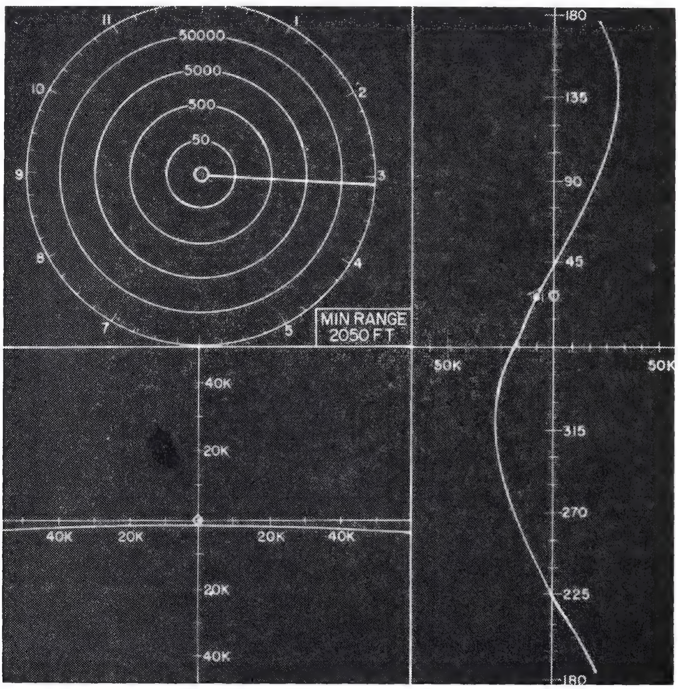


FIG. 34-5 *Transfer display.*

The instantaneous position of the destination satellite in its orbit is shown by a small circle that moves along the vertical line. The predicted path of the ferry vehicle for one complete orbit is shown, in terms of its deviation from the destination-satellite orbit, by means of a cathode-ray trace. If the ferry-vehicle transfer orbit and the destination-satellite orbit were coplanar, the trace showing the predicted path of the ferry would be coincident with the vertical line on the display. For orbital-plane differences causing deviations within the z scale on the display, the ferry's predicted-path trace would be a sinusoid varying about the η axis, as shown in Figure 34-5. The instantaneous position of the shuttle vehicle

relative to the target satellite is shown on the predicted-path trace by a small filled circle with an arrow pointing in the direction of vehicle travel.

A range-time presentation occupies a rectangle 5 in. high and 6 in. wide in the upper left corner of the display. The fixed display references are six concentric circles. The outermost circle contains 48 equally spaced index marks representing 15-min increments of time. The index marks at the conventional clock positions are labeled 1 through 12. The innermost

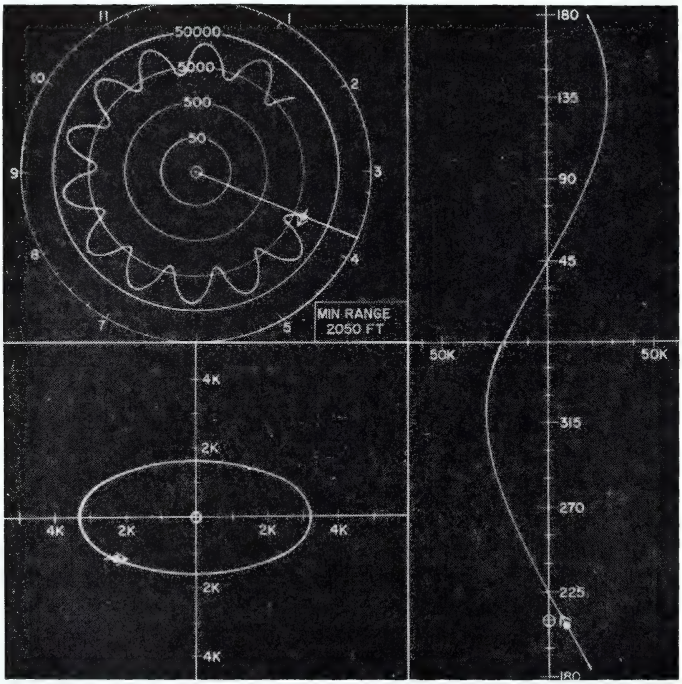


FIG. 34-6 *Rendezvous display.*

circle represents the destination satellite. The remaining four circles represent ranges from the destination satellite in feet on a logarithmic scale. The present time is shown by the position, relative to the time scale on the outer circle, of the end of a cathode-ray trace that appears as a radial line. The instantaneous range to the target is indicated by the position, relative to the range circles, of a dot (not shown in Figure 34-5) located on the radial line. Predicted range variations as a function of time for the next 12 hr may be shown as a cathode-ray trace emanating from the dot. A quantitative readout of the predicted minimum range is shown in the lower right corner of the *R-T* presentation.

Rendezvous Display

The rendezvous display, shown in Figure 34-6, is in only one way different from the transfer display. The display scale factor on the *X-Y* presentation in the rendezvous display is 2,000 ft/in., whereas the scale factor is 20,000 ft/in. on the *X-Y* presentation in the transfer display. The scale difference makes the rendezvous display appropriate for showing the position and predicted path of the ferry relative to the destination satellite during the terminal part of the orbit transfer, during closing maneuvers, and while the ferry is "parked" in the orbit of the destination satellite.

Figure 34-6 depicts a typical rendezvous situation. It is assumed that the orbit transfer has been completed by means of the second thrust application but that a closing maneuver has not been made. Present range to the space station is about 6,000 ft. The orbital elements of the two satellites are not matched exactly, but their orbital periods are approximately equal.

The *X-Y* presentation shows the predicted path of the ferry relative to the station, projected onto the orbital plane of the station for the next orbital period. As indicated by the predicted-path trace, while the two satellites are making one complete orbit the *x-y* motion of the ferry relative to the space station will describe an ellipse. The range-time (*R-T*) presentation shows the predicted range as a function of time for the next 10 hr.

Chapter 35

SYSTEM SIMULATION AS A TECHNIQUE IN SYSTEMS RESEARCH

*W. W. Haythorn**

IN THIS CHAPTER A NUMBER OF SYSTEM-SIMULATION EFFORTS IN WHICH human factors considerations play key roles are described with a view to illustrating techniques for examining human factors problems in a broad systems context. This kind of examination can assist in improving system performance in the following ways:

1. Provide better decision rules
2. Provide better tools (displays, controls, and equipment) for the human beings to manipulate
3. Help to automate functions normally performed by human beings
4. Provide better training of the men in the system
5. Provide a better information system to the decision makers

The importance of human factors in systems considerations is not new (see, for example, References 5, 8, 11, 15), but it has been only in recent years that techniques for coping with these considerations have been available. McGrath, Nordlie, and Vaughan [9] have recently reviewed systems research methodology; so it will not be necessary to do so here. The emphasis in this chapter will be on the application of simulation and modeling as techniques for examining human factors problems in systems. In this context, the construction of a simulation model is (1) a convenient form for recording what one knows and (2) a convenient form for examining the interactions among the facts that one has at his disposal in order to determine their compatibility with each other and with other known facts. In the research to be described the primary methodological device has been simulation combining symbolic representation of some aspects of systems, realized on a digital computer, with individual decision makers drawn from the population from which the ultimate users of the system are drawn [4].

* Logistics Department, The RAND Corporation, Santa Monica, Calif.

The work represented by this chapter was supported by Air Force Project RAND. The views expressed here are not necessarily those of The RAND Corporation.

SYSTEMS SIMULATION FOR TASK ANALYSIS

In their survey of factors affecting the degree of automation in test and checkout equipment, Swain and Wohl observed, "*There is no adequate systematic methodology in existence for allocating functions* (*in this case test and checkout functions*) *between man and machines*" [16]. However, on the basis of work performed by the author, it is suggested that simulation of a system would enable one to gain some clues as to such function allocation. For example, in Laboratory Project II (LP-II) of the RAND Logistics Systems Laboratory an extensive task analysis of the maintenance functions of ballistic-missiles systems was undertaken and the results built into a system simulation [7, 14]. Operating the simulated system under what were regarded as realistic conditions indicated that a much more efficient aggregation of tasks into positions could have been accomplished. The technique consisted of identifying qualitative personnel-requirements information through the usual analytic procedures: identifying equipment characteristics through an engineering analysis; aggregating tasks into positions; building a model that combined equipment, personnel, and policy data in essentially isomorphic relation to their real-world counterparts; subjecting the model to a likely range of real-world stresses; evaluating the outcome in terms of overall system effectiveness and cost; and iterating with a different aggregation of tasks into positions. The reaggregation was based on the frequency of the demands for particular skills, the similarity of skills [10], the frequencies of queues, and the percentage of effective utilization of personnel.

ESTIMATING MANPOWER REQUIREMENTS

In the development of a new weapon system the military support planners must make a variety of decisions based on assumptions regarding the personnel and equipment necessary to support the system. These decisions involve large dollar expenditures and typically must be made early in the development cycle because of the long lead time required for implementation. Consequently, the decisions must be made before the final design of the system has been determined and before the required information for supporting these decisions can be obtained.

The Logistics Systems Laboratory first began including manpower requirements in its simulation exercise in LP-I [13], in which they were represented simply as man-hours required to repair malfunctioning parts. All man-hours were equivalent. Manpower requirements were predicted by estimating the number of future repairs and multiplying by the man-hours required for each part to be repaired. The naïveté of this repre-

sentation was recognized, but for a study focused primarily on spare-parts procurement and distribution the procedure was adequate.

In LP-II, a systems simulation of a ballistic-missile organization [14], the problem was broadened to include personnel and equipment support. This required greater sophistication in the laboratory representation. Consequently, the personnel model in LP-II included the full range of Air Force Specialty Codes required for direct support, skill-level differences, and individual differences within skill levels. Skill-level differences were represented by tasks that the man could perform; individual differences were represented by differences in time required to perform the task. Tasks were represented by operational and maintenance situations for scheduled activity and by the intersection of component parts and maintenance jobs for malfunction-generated activity. For each task the personnel type, skill level, kind of equipment, and standard time required were specified. Each man in the unit had an aptitude index that determined the ratio of the time that it took him to do a job for which he was qualified to the standard time. This requirements-estimation technique will be described in the context of LP-II, but obviously it is generalizable to other hardware systems. (For more detail, see References 6, 7, and 12.)

An objective explicitly stated in many personnel-requirements estimates, growing out of job analyses, position descriptions, human engineering studies, and the like, is that of minimizing personnel requirements. In all aspects of LP-II, however, an attempt was made to increase the effectiveness-to-cost ratio, with effectiveness defined as the percentage of missiles that the system could normally launch in case of necessity. This objective was applied to manpower requirements, as it was to all other requirements.

An engineering analysis of ballistic-missile hardware was undertaken, which, among other things, developed malfunction rates for each type of equipment stress for each component part of the system [1]. For each component part the Air Force Specialty required to accomplish each of a number of maintenance tasks was also identified. This task-by-part matrix permitted the computation of the expected frequency of task occurrence for any constellation of equipment stresses. In estimating manpower requirements, two constellations of stresses were used, that for a normal, peaceful day, during which missiles were mostly standing on alert, with some time allotted to various kinds of maintenance activities, and that for a launch program, in which all available missiles were subject to countdown and launch stresses. For each of these two sets of stresses, a computation was accomplished that indicated the probability of occurrence of each of a possible 9,000 tasks. Summing these probabilities by Air Force Specialty Code produced the expected number of calls for each specialty during a normal working day and during a launch program.

The next question was how many of each of the different maintenance Air Force Specialty Codes should be assigned to the launch complexes and how many should be assigned to the central repair shops. The answer to this question was determined by adding men to the launch-enclosure maintenance crew until the addition of another man would cost more than the resulting increase in missile-launch capability was worth. This criterion stems from the fact that the system's objective was to maintain a launch capability and that, in the context of the present discussion, launch capability can be increased either by placing more maintenance capability at each launch complex or by buying more missiles.

Emphasis on the launch-enclosure maintenance crew was based on the assumption that missile systems are essentially quick-reaction weapons and that maintenance capability at the central support area would be useless after the beginning of any hostilities. Launch-enclosure maintenance crews made two distinct contributions to firing capability. First, they were available immediately in case of malfunctions, thereby saving the transportation time required for bringing men and equipment from a central area. Second, in case of a launch-program malfunction, they were on hand to take corrective action and prevent the missile's being completely useless.

The procedure, then, was to compute the percentage of missiles that one would expect to have on alert during normal operations if there were no maintenance men at the complex and if all necessary skills were available through the dispatch of specialists from the central support area. The difference between this situation and that of having all necessary skills available at the complex was simply that each call for unscheduled maintenance incurred a travel time in addition to the time required to correct the malfunction. The expected alert posture was then converted to an expected launch capability by multiplying the alert percentage by missile-launch reliability. For example, if the expected alert were 70 per cent and the missile reliability 50 per cent, the expected launch capability would be 35 per cent.

Next the computation was repeated, but with the one man assigned to the launch-enclosure maintenance crew who could make the greatest contribution to launching capability. The cost of adding the one-man crew, including the requirements for crew rotation, equipment, and additional facilities, was compared with the value of the increased launch capability, value being determined by what it would cost to get an equivalent increase by purchasing more missiles. This procedure was repeated, adding each time the one additional man who yielded the greatest gain until it was no longer cheaper to add men than to buy missiles. It should be noted that this is an insurance allocation procedure and in no way reflects the typical objective of minimizing manpower requirements or of maximizing manpower utilization.

Maintenance manning for the central support area was determined by estimating the total work load expected for the organization, including both scheduled and unscheduled maintenance, by assigning as much of it to the men already allocated to the launch complexes as they could be expected to perform (assuming that approximately 50 per cent of their available time could be spent in direct labor), and by adding the additional men required to meet the work load at the central maintenance area.

SYSTEM SIMULATION FOR TRAINING PURPOSES

In 1951, as part of a broader air-defense systems analysis, the simulation of manual Air Defense Direction Centers was undertaken. The initial purposes of these studies were to establish input data for a broader systems analysis and to study basic problems of organizational effectiveness and systems learning. Experienced operators from the Air Defense Command were assigned positions in the simulated Air Defense Direction Center. This simulated center was as nearly a one-for-one replica of the real thing as the experimental staff was able to achieve. The experimental strategy was, following a period of training, to subject the crew systematically to air situations of progressively increasing stress, i.e., with progressively more tracks to identify and intercept if necessary. These studies are described in some detail by Chapman et al. [3] and by Sweetland and Haythorn [18].

The most important result of these studies was that the crews learned to cope with progressively more difficult situations until they were handling air-traffic loads considerably beyond what anyone had initially considered possible. This seems to have resulted from the application of a few simple principles, which can be summarized as follows.

1. Sufficiently realistic and challenging tasks must be presented.
2. In order to promote effective system or team training, the entire team must be trained as a unit.
3. Objective knowledge of the team's performance must be provided.
4. The difficulty of the task must be matched to the current ability of the team.

A similar strategy was applied in LP-II of the Logistics Systems Laboratory [14]. In this project Air Force officers drawn from the population from which real-world missile-squadron officers were being drawn were required to manage a simulated ballistic-missile organization. The detail included in the simulation was determined by an equipment analysis and by laboratory-staff constraints. Stress on the system was changed progressively by increasing the size of the organization to be

managed and by decreasing the amount of resources available to fulfill the task, thereby requiring tighter management. The most striking feature of LP-II, from a human factors point of view, was the high involvement of participants in the task and the ingenuity shown in solving the management problems. The solutions have been described in detail by Sweetland [17] and in more summary fashion by Haythorn [7].

DECISION ANALYSIS IN SYSTEMS

One of the key roles played by men in complex systems is that of a decision maker, a decision being defined crudely as the selection of a response from a pool of potential responses. In a system as complicated as an Air Force ballistic-missile organization, the parameters relevant to system performance are seldom thoroughly understood, and their interrelations are never completely understood. This fact makes an analytic approach difficult and crude.

In an analysis of the surveillance functions of an Air Defense Direction Center, Sweetland and Haythorn described a dichotomous decision model that at least appeared consistent with a great host of empirical data [18]. This model is reproduced in Figure 35-1. The model indicates

Stimulus track

a Detected ? (seen) → no — no report
↓
Yes
↓
b Is it an established track ? → yes — report it
↓
No
↓
c Is it potentially critical ? → no — forget it
↓
Yes
↓
d Are we already carrying our limit ? → no — report it
↓
Yes
↓
e Can we drop something less important ? → no — delay actior
↓
Yes
↓
f Drop old track and report new

FIG. 35-1 *A dichotomous decision model of the surveillance function of an Air Defense Direction Center. (After A. Sweetland and W. W. Haythorn, An Analysis of the Decision-making Functions of a Simulated Air Defense Direction Center, Behavioral Sci., vol. 6, pp. 105–116, 1961. By permission of the publishers.)*

a number of decisions made by a scope operator in conjunction with a movements-board plotter in simply determining whether or not a blip on a radar screen will be reported as an aircraft track. The model points to a number of places where improvement could be achieved, through modifying procedures, through improving tools, or through changing the organization. The learning that occurred in the simulated air-defense system was found primarily in step *e* of the model. This learning resulted in an earlier dropping of already identified, and therefore less important, tracks in order to pick up and carry newly found blips. The conclusion drawn from this analysis was that in dealing with their essentially information-transferring function the operators of the system had an intuitive model of reality that they compared with the information that they acquired through their radars. This model was analyzed, with the results determining the responses of the system. Throughout the system there appeared to be a normative activity level within which the operators tried to function. If the activity level required for dealing with the immediate situation increased, operators brought it back into line by changing the rules regarding which tracks they would carry and which they would not. As the activity level indicated by the stimulus dropped below a comfortable activity level, operators carried tracks that they normally would not. In fact, it has been noticed that, as the activity level required by the stimulus becomes too low, operators invent tracks and carry them apparently just for the fun of it.

In Laboratory Project II, Sweetland analyzed the decision process of maintenance and operations officers in scheduling the activity of their simulated ballistic missile program [17]. He found that, as more and more sophisticated scheduling techniques were developed by the participants, the officers became increasingly able to improve the operational capability of their system at the same time as they were able to reduce the manpower and equipment required to meet their objectives. This decision analysis resulted in a considerable improvement in the effectiveness-to-cost ratio, which was the primary criterion of system effectiveness. These decision processes, developed and observed in a laboratory context, appear to have real-world relevance.

PARTICIPANT FEEDBACK ON SYSTEMS DESIGN

The design and development of a modern weapon system is a long, drawn-out affair involving a large number of organizations and subgroups within organizations. A number of developmental activities must proceed parallel to and more or less independently of each other. This results inevitably in a certain amount of incompatibility among plans

and may lead to a costly and time-consuming patchwork and retrofit program before the system becomes fully operational. In LP-II and LP-III (a study of an advanced set of repair and distribution policies implemented with the aid of electronic data-processing equipment), a great deal of use was made of observations and comments by the participants in the identification and improvement of defects in the initial system design. In both studies, men experienced in the real-world functions for which the systems were being designed were given key management roles in the operation of the systems. By being thoroughly immersed in the details and highly involved in the outcome, these participants were able to provide a great deal of feedback into the design process. It has been suggested that this use of simulation might bring the reactions of the ultimate customers into the design process early enough to affect a system's final configuration and method of operation.

SIMULATION FOR RESEARCH IN HUMAN FACTORS

In an article by Chapman the value of simulation for studying human behavior is discussed [2]. The emphasis on providing individuals with a realistic and challenging environment in a laboratory setting is well placed. The advantages of laboratory studies over real-world observation are well known. Laboratories provide greater control capability and greater access to the necessary data, and they are considerably less expensive and less risky than real-world manipulation studies. In the simulated Air Defense Direction Centers, for example, an opportunity was provided for studying the process by which organizations adapt to increasingly difficult situations. The freedom to make changes and to evaluate these changes in a system's context was believed to be largely responsible for a remarkable increase in organizational effectiveness. Some advances were made in developing a theory of system adaptation. It is difficult to see how such research can be undertaken through any but a simulated systems-research laboratory approach.

BIBLIOGRAPHY

1. Bean, Eloise E.: A Malfunction-generation Model for a Manned Simulation of ICBM Logistics, *RAND Corp. Paper* P-2317, May 22, 1961.

2. Chapman, R.: Simulation and Human Behavior, *Am. Management Assoc. Management Rept.* 55, pp. 91–108, 1961.

3. ——— et al.: The Systems Research Laboratory's Air Defense Experiments, *Management Sci.*, vol. 5, pp. 250–269, April, 1959.

4. Geisler, M. A.: The Use of Monte Carlo Models, Man-Machine Simulation, and Analytical Methods for Studying Large Human Organizations, *RAND Corp. Paper* P-1634, Mar. 11, 1959.

5. Goldhamer, H.: Human Factors in Systems Analysis, *Rand Corp. Research Mem.* RM-388, ASTIA ATI 78026, Apr. 15, 1950.

6. Haythorn, W. W.: Maintenance Scheduling Decisions and the Importance of Information, *RAND Corp. Paper* P-2302, May 9, 1961.

7. ———: Simulation in Logistics Policy Research, *RAND Corp. Paper* P-1791, Sept. 10, 1959, in P. Rosenstiehl and A. Ghouila-Houri, "Les Choix économiques, décisions sequentielles et simulation," pp. 211–236, Dunod, Paris, 1960.

8. Javitz, A. E. (ed.): Engineering Psychology and Human Factors in Design, *Electro-Technol.*, vol. 29, pp. 108–130, May, 1961.

9. McGrath, J. E., P. G. Nordlie, and W. S. Vaughan, Jr.: A Systematic Framework for Comparison of System Research Methods, *Human Sci. Research Tech. Note* 59/7-SM, November, 1959.

10. Miller, R. B., J. D. Folley, Jr., and P. R. Smith: A Comparison of Job Requirements for Line Maintenance of Two Sets of Electronics Equipment, *USAF Personnel Train. Research Center Tech. Rept.* AFPTRC-TR-54-83, December, 1954.

11. National Research Council, Committee on Undersea Warfare, Panel on Psychology and Physiology: "Human Factors in Undersea Warfare," National Research Council, Washington, 1949.

12. Pickrel, E. W.: Support Resources, *RAND Corp. Paper* P-1981, May 10, 1960.

13. Rauner, R. M.: Laboratory Evaluation of Supply and Procurement Policies: The First Experiment of the Logistics Systems Laboratory, *RAND Corp. Rept.* R-323, Aug. 20, 1958.

14. ——— and W. A. Steger: Simulation and Long-range Planning for Resource Allocation, *RAND Corp. Paper* P-2223-1, Mar. 15, 1961.

15. Rosenstiehl, P., and A. Ghouila-Houri: "Les Choix économiques, décisions sequentielles et simulation," Dunod, Paris, 1960.

16. Swain, A. D., and J. G. Wohl: Factors Affecting Degree of Automation in Test and Check-out Equipment, *Dunlap & Associates Tech. Rept.* 60-36F, Mar. 1, 1961.

17. Sweetland, A.: Operations and Support Scheduling Methods Derived from Laboratory Problem II (LP-II)—A Manned ICBM Simulation, *RAND Corp. Research Mem.* RM-2669, Nov. 15, 1960.

18. ——— and W. W. Haythorn: An Analysis of the Decision-making Functions of a Simulated Air Defense Direction Center, *Behavioral Sci.*, vol. 6, pp. 105–116, 1961.

Chapter 36

THE RELATION OF BLUR AND GRAIN TO THE UPPER LIMIT OF USEFUL MAGNIFICATION

*Glenn A. Fry**

THERE ARE MANY FACTORS THAT HELP TO SET THE UPPER LIMIT TO THE amount of magnification that can be used to good advantage in viewing a photograph. Blur and grain in the photographic image are both affected by magnification. Consideration will be given here to the effect of blur and grain on the visibility of lines, points, and borders at high levels of magnification. The special problems connected with the resolution of gratings have been dealt with elsewhere [5].

Magnification will harmfully affect the visibility of a long, straight border because it will increase the blur and eventually make the grains visible. Once the grains are visible, magnification increases their size and the coarseness of their spacing. Both the blur and the grain impair visibility. Magnification also increases the length of the border; however, if the border is already long at the outset, the increase in length per se will not affect the visibility and therefore cannot compensate the loss from blur and grain. On the other hand, magnification does, up to a point, increase the visibility of curved and segmented borders, lines, points and small disks, and rectangles because other factors more than compensate the effects of blur and grain.

VISIBILITY OF LONG, STRAIGHT BORDERS

A contrast border in an ordinary photograph viewed at the ordinary viewing distance without a magnifying device appears sharp and is con-

* School of Optometry, Ohio State University, Columbus, Ohio.

The research reported in this chapter was supported by a contract between the Air Development and Research Command, U.S. Air Force, Rome Air Development Center, and Ohio State University Research Foundation.

The author is indebted to Dr. Gordon Bixel, Jr., and to Miss Mildred Hindman for their assistance in the execution of this study.

strued as an abrupt transition from one luminance level to another. This kind of border is illustrated in Figure 36-1.

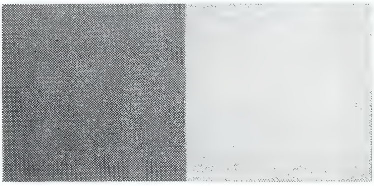


FIG. 36-1 *Appearance of a photographic image of a straight border viewed without a magnifying device.*

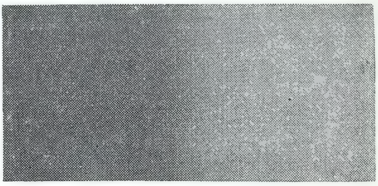


FIG. 36-2 *Appearance of a photographic image of a straight border magnified enough to show the blur but not enough to show the grain.*

If the grain in the emulsion is fine, so that the photograph may be magnified many times before the grain becomes visible, there is an intermediate level of magnification at which the border is perceived as a gradual transition from one level of luminance to another, as illustrated in Figure 36-2.

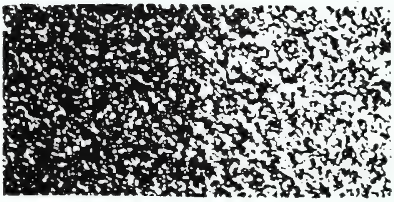


FIG. 36-3 *Grainy pattern used to simulate the appearance of the photographic image of a straight border magnified enough to show the grain.*

At some level of magnification the grain becomes visible and the smooth gradient between two uniform patches is transformed into a grainy gradient between two grainy patches, as illustrated in Figure 36-3.

As long as the magnification is low enough so that no grain is visible, a border may be regarded as a smooth gradient between two levels of luminance, B_{left} and B_{right}, as illustrated in Figure 36-2. The gradients studied have been precisely generated and may be represented by a curve (Figure 36-4) showing B as a function of x,

$$B = B_L + (B_R - B_L)\frac{1}{\phi\sqrt{\pi}}\int_{-\infty}^{x} e^{-(x/\phi)^2}\,dx \tag{36-1}$$

where x represents distance from the center of the gradient and where ϕ is the Fry-Cobb index of blur [2]. Both x and ϕ are expressed in minutes of visual angle. Contrast in this case may be defined as follows:

$$\text{Contrast} = \frac{B_R - B_L}{B_L} \tag{36-2}$$

The apparatus for producing this type of stimulus pattern is described in detail in another paper by the author [3]. It is sufficient to say that the edge of the photograph in Figure 36-2 is the edge of a rectangular aperture (1 by 2 in.) in a white screen viewed monocularly at a distance of 1 m through a 2-mm artificial pupil and through a lens that permits the subject's accommodation to be relaxed. Filling the aperture is an aerial image of a bipartite field with a blurred contrast border as a dividing line between the right and left halves of the aperture. The amount of blur can be changed by varying the magnification of this image. The index of blur (ϕ) is directly proportional to the magnification. The contrast can be varied independently. The luminance of the

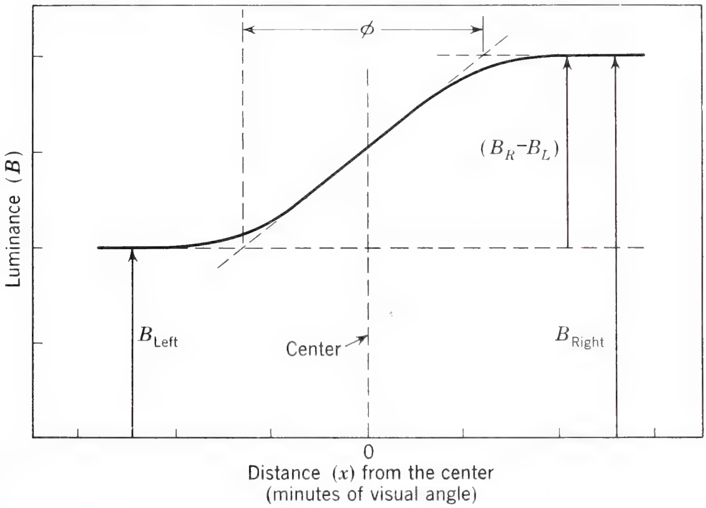


FIG. 36-4 *Method of specifying blur and contrast.*

surround was constant at 33 ft-lamberts, and the luminance on the right side of the border (B_R) was kept constant at 23 ft-lamberts. The luminance on the left side (B_L) was varied to change the contrast.

Figure 36-5 shows the type of data obtained in this kind of experiment. In the case of the solid circles the contrast was held constant while a series of threshold measurements was made by varying blur. The criterion was the ability to distinguish an abrupt border. In the case of the open circles the blur was kept constant while a series of threshold measurements was made by varying contrast. Each point is an average of five settings. At each of the lower levels of blur two thresholds were found, a lower one, at which the brightness difference between the two halves was detected, and a higher one, at which the gradient became transformed perceptually into an abrupt border. At high levels of blur,

one does not encounter an abrupt border even at the highest levels of contrast; therefore only the threshold for a perceptible difference in brightness can be reported.

It is clear from the results that one can make a judgment of the brightness difference even though there is no abrupt transition at the border. The subjects were instructed not to report a brightness difference simply on the basis of a brightness difference at the two extreme edges of the rectangle; rather, they were to report a difference when they could see a border or gradient between two uniform areas, one on each side.

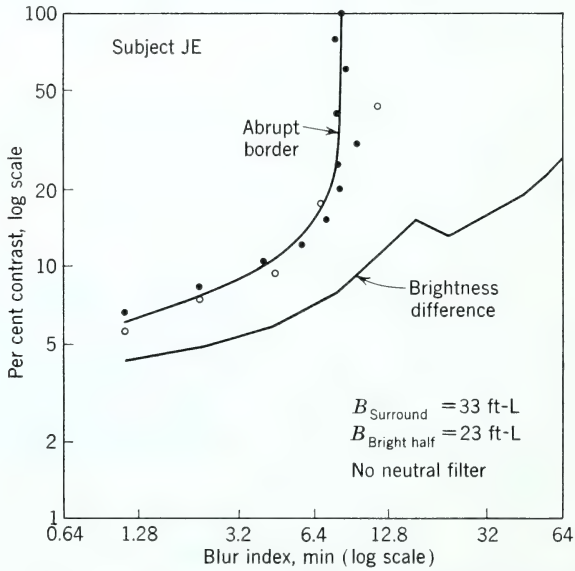


FIG. 36-5 *The effect of blur and contrast on the threshold of visibility and the perceived abruptness of a border.*

Small amounts of blur produce a negligible effect upon the contrast required for a just-perceptible difference in brightness, but as the blur reaches larger degrees the threshold begins to rise and continues to rise.

It is interesting to compare a different situation, in which the grain size is large in comparison with the amount of blur and, as a result, the grain becomes visible when the amount of blur is gradually increased. Patterns like that in Figure 36-3 have been used to study this effect. Figure 36-3 shows a grainy blurred border separating the right and left halves of the rectangle. The technique for producing these patterns has been described elsewhere [5]. This figure illustrates that one can assign two levels of brightness to the two halves of the pattern in spite of the fact that the pattern is composed of discrete black and white areas. The

question that must be answered is whether the discrimination between the two halves of the target is based upon the difference in perceived brightness or upon some subtle difference in the shape and size of the black and white areas on the two sides of the border.

What is meant by discrimination based on a difference in size and shape of the black and white areas on the two sides of a border is illustrated in Figure 36-6. Here the two sides of the border have the same average luminance, and if the pattern is viewed from a sufficient distance, there is no perceptible difference in brightness. Hence discrimination between the two halves of the pattern at the ordinary reading distance has to be based on the fact that the dots are black on one side and white on the other. There is also a definite shore line that divides the black field peppered with white dots from the white field peppered with black dots.

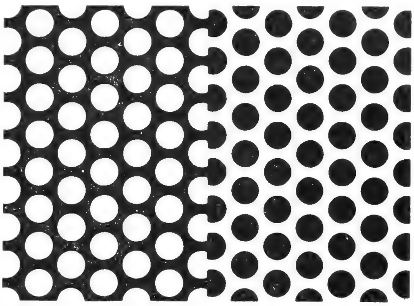


FIG. 36-6 *Pattern having the same average luminance but noticeably different grain pattern on the two sides of the border.*

In Figure 36-3 a shore line can also be traced, but this line is tortuous and is not directly perceptible. Except for the brightness difference between the two halves, about all that one can tell from direct observation is that the black grains appear to be smaller and more thinly distributed on the right. As long as one perceives the black areas on the right as grains, one is likely to perceive the white areas on the left as openings between grains and is not, therefore, aware of a transition between white grains on the right and black grains on the left.

The pattern in Figure 36-7 is made up of white grains that differ merely in average size and spacing. In Figure 36-8 both sides of the

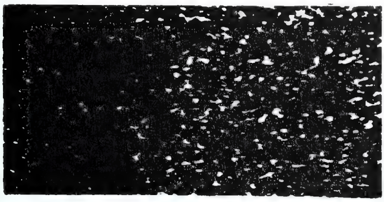


FIG. 36-7 *Blurred contrast border produced with white grains of different average size and spacing on the two sides.*

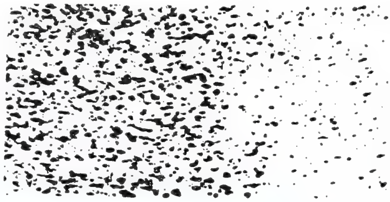


FIG. 36-8 *Blurred contrast border produced with black grains of different average size and spacing on the two sides.*

pattern are made up of black grains that differ in average size and spacing. In both these patterns the presence of the border has to be detected as a perceptible brightness difference and cannot be detected by the pattern formed by the grain. This can be demonstrated by combining each of these patterns with one like that in Figure 36-2, to produce a combination having a uniform average luminance from one side of the target to the other.

Not only is it possible to eliminate the perceived brightness difference, but the difference in average grain size and spacing is no longer visible at a casual glance. Wholly unexpected, this result has an important implication: in the typical photographic image the difference in grain pattern on the two sides of a border is of no value in detecting the presence of a border; there has to be a difference in average luminance. The only possible exception to this rule is the case in which there are grains on one side of a border but no grains on the other, as in Figure 36-9. At high magnification the grains are visible as grains, but the brightness difference on the two sides of the border is not visible at low magnification. Combining a target like that of Figure 36-7 with one like that of Figure 36-2 makes it possible to demonstrate that an increase in magnification beyond the point at which the grain is visible impairs the ability of the eye to discriminate a brightness difference.

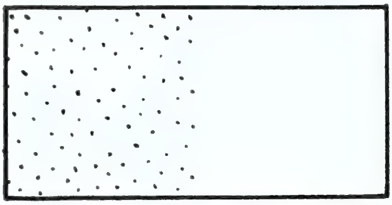


FIG. 36-9 *Target in which the grain pattern in the right half can be detected at high magnification but the right half cannot be discriminated from the left half at low levels of magnification.*

A pattern of the type shown in Figure 36-7 was made up as a transparency and scanned with a slit parallel to the grainy gradient in order to determine the average transmittance at various distances from the center of the gradient. From the resulting data the ϕ value of the pattern in terms of distance across the transparency was found to be 12.5 mm. The pattern was then mounted in front of a bipartite field, similar to the one shown in Figure 36-2, with a blurred dividing line having a ϕ value of 4.6 mm. The dark half of the grainy pattern covered the bright half of the bipartite field. The contrast between the two halves of the bipartite field could be varied. This combination was observed by forming an aerial image of it in a 2-in. by 1-in. rectangular aperture 1 m from the eye. It was viewed monocularly through a 2-mm artificial pupil and through a lens that permitted the subject to relax accommodation. The average luminance on the left side was kept constant at 4.1 ft-lamberts, and that on the right was varied to determine the two limits, the one at

which the right half was perceived as darker and the other at which it was perceived as lighter. Even though the ϕ values for the two components were unequal, there were no brightness variations at the border when the two halves were perceived as equal. The magnification of the aerial image was varied and is expressed in Figure 36-10 in terms of minutes of visual angle per millimeter in the plane of the transparency.

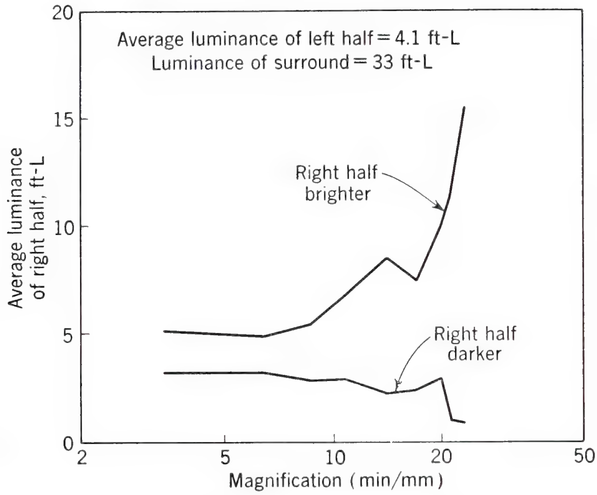


FIG. 36-10 *Effect of magnification above the level at which grain is visible upon the perception of a brightness difference on the two sides of a blurred border.*

The grain spacing was assessed for each half of Figure 36-7 by determining the average number of black segments per millimeter along a straight line, and the grain size was assessed by determining the average length of the black segments. The results of these measurements were as follows:

	Bright half	Dark half
Diameter of dark grain, mm.....	0.19	0.61
Grains/mm.....................	2.60	1.35

At the smallest magnification used, the grain was still visible, but not markedly so.

At various magnifications the subject determined the limits at which the right half was perceived as different in brightness from the left half. The data for one subject are shown in Figure 36-10. Each point is an average of five settings. It is clear that as the magnification increases the ability to detect a difference in brightness decreases. This effect may be due in part to the increase in blur, but it may also be dependent upon

the increase in the difficulty of assessing the average brightness in the two halves.

This question was investigated by using the polka-dot pattern in Figure 36-11 in combination with the pattern in Figure 36-1 in the same way as the combination of the patterns in Figures 36-2 and 36-7 was used. The blur was reduced to almost zero. Since there is no difference in grain size, spacing, or configuration between the two sides of the border, any effect of magnification upon the perception of a brightness difference must be attributed to the increased difficulty of assessing average brightness with increased grain spacing. The just-perceptible difference in brightness was measured at various magnifications. For this pattern the contrast has been defined as follows:

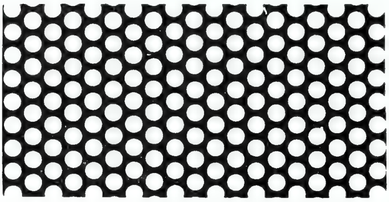


FIG. 36-11 *Polka-dot pattern.*

$$\text{Contrast} = \frac{(\text{av. luminance of left half}) - (\text{av. luminance of right half})}{\text{av. luminance of left half}}$$

The results are shown in Figure 36-12. The grain becomes visible at a spacing of about 2 min between rows. As the magnification increases

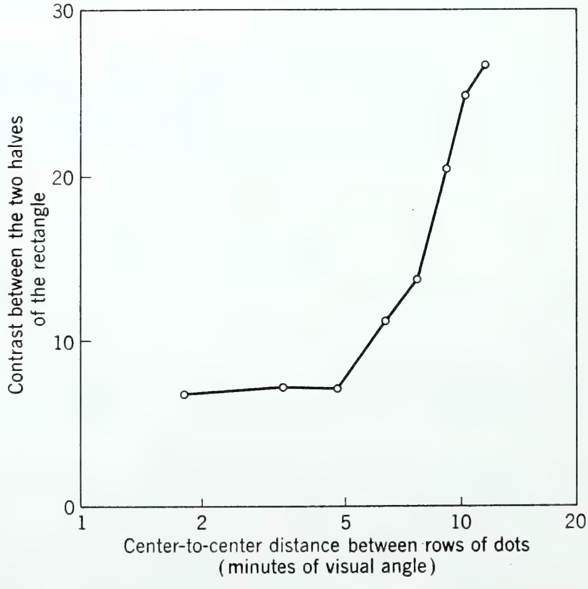


FIG. 36-12 *Effect of magnification above the level at which grain is visible upon the perception of a brightness difference on the two sides of a sharp border.*

there is also an increase in average luminance required for a perceptible brightness difference. Since the effect cannot be attributed to an increase in blur, it is demonstrated that visible grain does interfere with the perception of a brightness difference.

FIG. 36-13 *Enlarged half-tone image of a left eye.*

The pattern in Figure 36-13 illustrates how magnification of a complex grainy pattern can interfere with the extraction of meaning through the loss of brightness relations. If the pattern is reduced in magnification, as shown in Figure 36-14, the impression of an eye becomes more apparent. Furthermore, blurring the pattern in Figure 36-13 by placing a ground glass a short distance in front of it will restore the brightness differences and make the impression of the eye emerge.

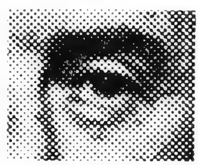


FIG. 36-14 *A reduction in size of the picture in Figure 36-13.*

VISIBILITY OF POINTS AND SMALL DISKS

The photographic image of a point is not a point, but a blurred spot graded from the center out to where it becomes indistinguishable from the background. The curve in Figure 36-15*a* represents the distribution of luminance across the center of the image of a dark point on a bright background.

What is the effect of magnifying such an image? The answer to this question could be obtained by direct experiment, but it appears that up to now no such experiment has been performed. It is possible to obtain an approximate answer to this question by an indirect approach. When the image is small, its visibility can be predicted from data for sharply defined disks; when the image is large, its visibility can be predicted from data for a blurred straight border.

It is necessary to assume certain relations between the blurred image of the point, the small disk, and the blurred border that are to be compared. These relations are shown in Figure 36-15. Figure 36-15*a* is

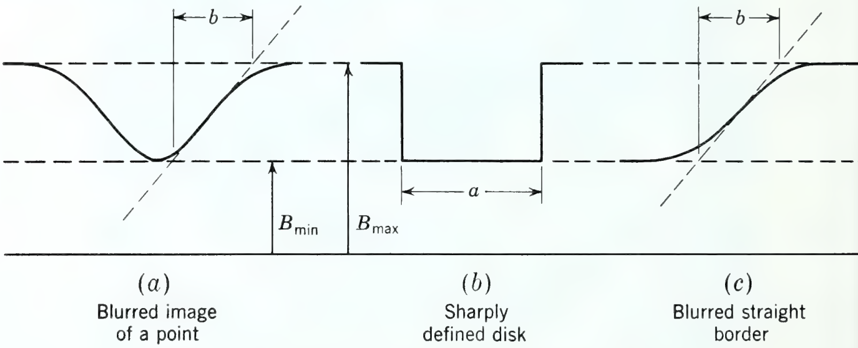


FIG. 36-15 *Distribution of luminance across the centers of three types of stimulus patterns. In each case the contrast is equal to* $(B_{max} - B_{min})/B_{max}$.

the distribution of luminance across the blurred image of a point. Since this is an inverted normal curve, the Fry-Cobb index of blur, ϕ, is related to the dimension b as follows:

$$\phi = 1.52b$$

Figure 36-15*b* is the distribution of luminance across the disk. The volume of the disk image

$$\pi \left(\frac{a}{2}\right)^2 (B_{max} - B_{min})$$

is equal to the volume of the blurred image of the point

$$2\pi \int_0^\infty (B_{max} - B)r\, dr$$

where r represents distance from the center of the image. The dimension a is 1.72 times the dimension b. Figure 36-15*c* is the distribution

of luminance across the blurred straight border. The dimension b for this distribution is the same as the dimension b for the distribution in Figure 36-15*a*.

Data for a sharply defined disk viewed at different levels of magnification were obtained with the stimulus pattern shown in Figure 36-16. The brightness of the surround was 33 ft-lamberts, and that of the immediate background ($B_{\max}$) was 23 ft-lamberts. The open circles in Figure 36-17 represent the threshold values of contrast for different values of a.

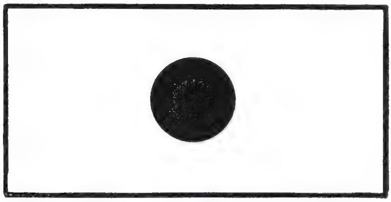


FIG. 36-16 *Stimulus pattern used to measure the contrast threshold for a small disk.*

Data for a blurred straight border viewed at different levels of magnification were obtained with the stimulus pattern shown in Figure 36-2. The brightness of the surround was 33 ft-lamberts; the brightness $B_{\max}$ on the right side of the border was constant at 23 ft-lamberts. The solid circles in Figure 36-17 represent the contrast threshold for various values of b. These threshold measurements were made in the same way as those for the lower curve in Figure 36-5. Both sets of data shown in Figure 36-17 were obtained with the same subject.

The contrast thresholds for the blurred image of a point (Figure 36-15*a*) can be predicted and can also be plotted in Figure 36-17, for the scales for a and b are arranged so that, as the magnification increases, a and b maintain a constant ratio 1.72, which makes the volume of the blurred image of the point equal to the volume of the image of the disk. It should be noted that the curve for the sharply defined disk levels off at a contrast of about 2.5 per cent as the size increases and that this conforms to the value for the straight edge at small values of b. It may therefore be predicted that the curve for the blurred image of a point will become tangent to the curve for a blurred straight edge at an a value of about 60 minutes of arc. Also, at small values of a, the blurred image of a point as well as the clear image of a small disk must conform to Ricco's law; i.e., the two curves must coincide and must both have a slope of -2 [4].

At some intermediate level of magnification the threshold value for the blurred image of a point reaches a minimum; this value represents the upper limit of useful magnification. Let us assume for the moment that this occurs at a b value of 16.5 minutes, which corresponds to a ϕ value of about 25 minutes. The amount of magnification required to reach the upper limit will depend upon the ϕ value of the blurred image of a point before it is magnified. The original value of ϕ will depend not

only upon the blur index for the camera but also upon the gamma value of the film, the spread of the image in the film, and the length of the exposure and development processes. In general, the following rule may be made: the upper limit of useful magnification is equal to the number 25 divided by the blur index of the photographic image expressed in minutes of visual angle.

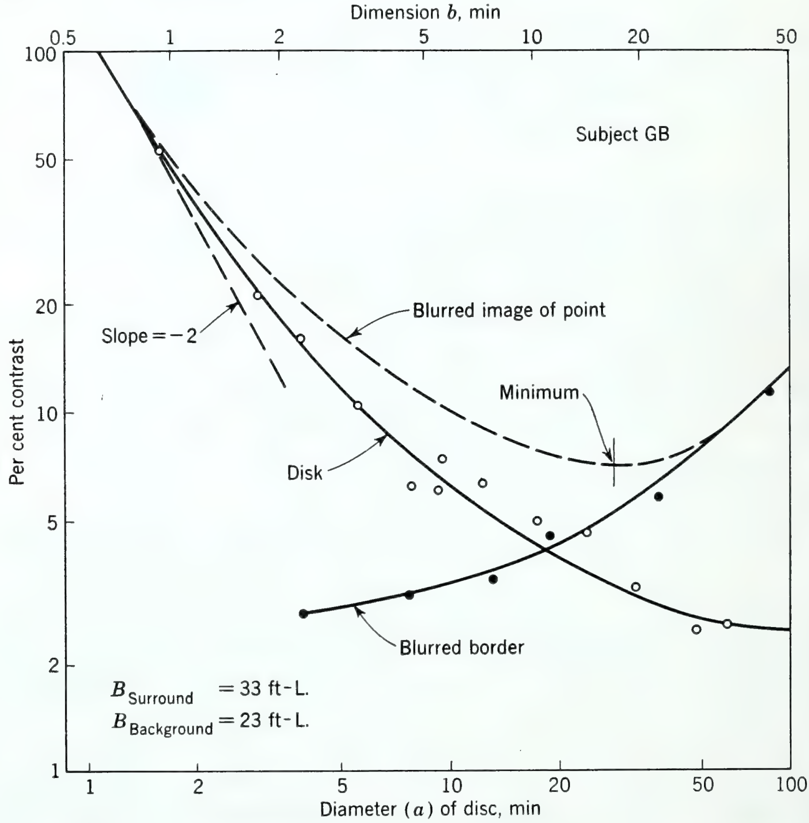


FIG. 36-17 *The relation of the contrast thresholds for a blurred image of a point to those for a blurred straight border and those for a sharply defined disk viewed at various levels of magnification.*

If a blurred image of a disk is used instead of the blurred image of a point, the same principles will apply in determining the upper limit of useful magnification but the amount of magnification required for reaching the upper limit will decrease as the size of the disk approaches the limiting diameter of 60 minutes. Enlarging a disk beyond 60 minutes will not improve its visibility even when the border is sharply defined. If the grain becomes visible before the theoretical upper limit of useful

magnification is reached, the grain will lower the upper limit. The combined effect of grain and blur has yet to be determined.

THE VISIBILITY OF A RECTANGLE

The threshold for a square varies with size in much the same way as for a disk. However, whereas the threshold for a narrow rectangle is affected by varying the length up to about 60 minutes or more [4], it is not affected by changing the width except for very narrow widths up to about 3 to 5 minutes of visual angle. Hence the effect of magnification is not nearly so dramatic for a line or narrow bar as for a small square.

FIG. 36-18 *Serrated border.*

FIG. 36-19 *Wavy border.*

That the effect of varying the width of a narrow bar upon its contrast threshold extends only up to about 3 to 5 minutes of visual angle implies that this effect can be attributed to blur of the ocular image and to physiological irradiation. Some other mechanism must be sought to account for the effect of varying the length of a rectangle or the diameter of a disk that extends up to 60 minutes of arc. In the case of the square or rectangle, the length of the sides appears to be the important variable, rather than the area; in the case of the disk, the significant factor is the length and curvature of the border. This can be demonstrated by using serrated and wavy borders, such as those shown in Figures 36-18 and 36-19, where area is not affected by magnifying the serrated or wavy border [4]. Data obtained with patterns of this type are presented in Figure 36-20. At low levels of magnification, two thresholds are found, one at which a straight border is perceived and one, the higher threshold, at which the detail becomes visible. At a certain level of contrast, wavy and serrated borders are visible as borders but cannot be differentiated from each other. This illustrates the general way in which image degradation gradually reduces the amount of information that is transmitted. It is easy to see how blur or grain would degrade the visibility of wavy and serrated borders in much the same way as reduced contrast would.

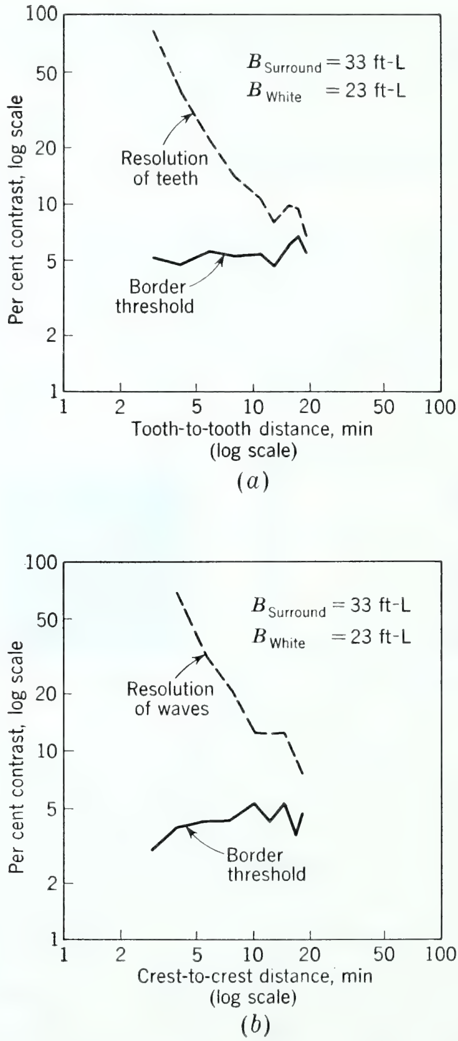


FIG. 36-20 *Data for one subject for a serrated and for a scalloped border.*

Barrows [1] has studied the effect of blur and grain upon the differentiation of squares and disks.

COMPLEX PATTERNS

The specific stimulus patterns that have been considered up to this point represent a few of the major types of simple configurations. Besides many other configurations, there are complex combinations of adjacent,

adjoining, and crossing borders that have still to be considered. A thorough, systematic analysis of this whole problem has yet to be made. There is little reason to believe that any of these patterns would benefit from any more magnification than the image of a point. Thus it is recommended that the image of a point be used as the basis for setting the upper limit of useful magnification.

BIBLIOGRAPHY

1. Barrows, R. S.: Factors Affecting the Recognition of Small Low-contrast Photographic Images, *Phot. Sci. Eng.*, vol. 1, no. 1, pp. 15–22, 1957.

2. Fry, G. A.: "Blur of the Retinal Image," Ohio State University Press, Columbus, Ohio, 1955.

3. ———: The Relation of Blur to the Visibility of a Brightness Contrast Border, unpublished report, 1961.

4. ———: The Relation of the Configuration of a Brightness Contrast Border to Its Visibility, *J. Opt. Soc. Am.*, vol. 37, pp. 166–175, 1947.

5. ——— and J. M. Enoch: The Relation of Blur and Grain to the Visibility of Contrast Borders and Gratings, *Ohio State Research Foundation Mapping Charting Research Lab. Tech. Paper* (606)-19-292, Contract AF 39(602)-1580, May, 1959.

Chapter 37

THE MEASUREMENT OF THREE-DIMENSIONAL HUMAN MOTIONS

*Gerald Nadler**

THE STUDY OF WORK METHODS REQUIRES DETAILED KNOWLEDGE OF MANY aspects of body-member motions: what a body member does when it positions or assembles an object, the effect of one motion on another in structuring a motion pattern, the exact motion differences between various methods suggested for the same job, what makes one operator's method different from another's on the same job, what constitutes a complete description of a method used by an operator, the techniques for determining regularity or changes in method, the characteristics of motion performance that can be used to measure operator-performance level, the effect of various components of difficulty on the performance of motions and motion patterns, and the nature of fatigue and its effect on the performance of motion.

Motion is measurable in terms of acceleration, velocity, deceleration, position in space, true distance traveled, and time. When such measurements are made for body members, certain factors have to be considered. The measurements must be three-dimensional. Any attachments to the body member studied must be small and light. Any measurement device developed must be usable under actual operator conditions. The range of velocities might be rather small (0 to 10 ft/sec), whereas the range of acceleration and deceleration could be large.

After investigation of many possible approaches for making these measurements [4], it appeared that the Doppler effect with sound as the radiation medium presented the fewest obstacles to the successful development of a work-measuring device. The device that was developed

* Department of Industrial Engineering, Washington University, St. Louis, Mo.

This chapter is derived from three technical reports: Donald Franz and Gerald Nadler, New Measurements to Determine the Effect of Task Factors on Body Member Acceleration Patterns, *J. Ind. Eng.*, vol. 12, no. 5, pp. 317–323, September–October, 1961; Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-Plane Motion Learning, *J. Ind. Eng.*, vol. 9, no. 3, pp. 187–197, May–June, 1958; Gerald Nadler and Jay Goldman, The Unopar, *J. Ind. Eng.*, vol. 9, no. 1, pp. 58–65, January–February, 1958.

is called the Universal Operator Performance Analyzer and Recorder (UNOPAR) [8].

The UNOPAR operation, diagramed in Figure 37-1, begins with an inaudible sound source (20,000 cps) emitted from a small (¾-in.-diameter, $\frac{1}{16}$-in.-thick) transducer attached to the body member of

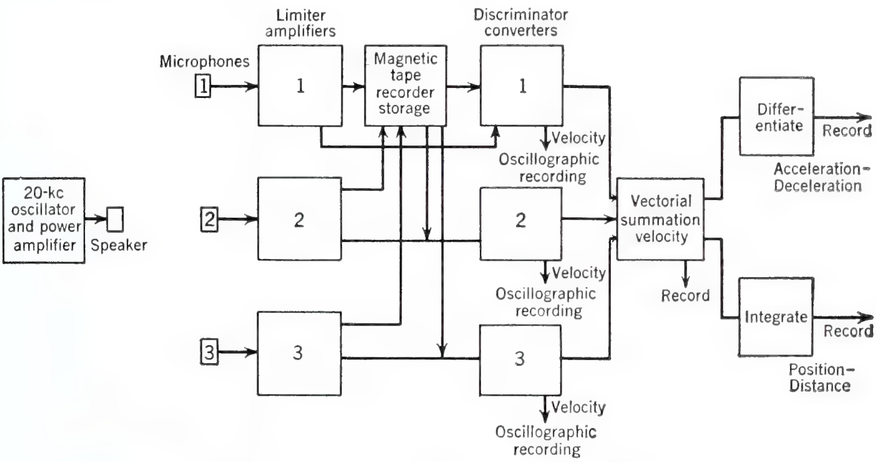


FIG. 37-1 *Block diagram of the UNOPAR operation.*

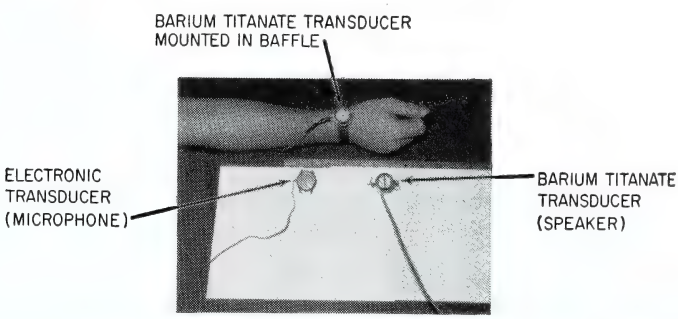


FIG. 37-2 *The transducers used for emitting 20 kc and receiving the signal from the moving body member.*

interest, as shown in Figure 37-2. If the motions of more than one body member or more than one point on a body member are to be studied, different emission frequencies can be used.

The emitted sound is received by three microphones, which are situated along mutually perpendicular axes around the work area. If each microphone is located about 10 ft from the subject, motions within a 1-yd cube maintain the geometric accuracy of the coordinate system to within ±1 per cent. Because almost all work activity is performed

within a 1-yd cube, the size of the cube and the distance of the microphones from the cube can be increased, and the cube may be oriented in any manner to encompass most of the motions in an operation, this error problem is insignificant for practical purposes.

The signal to each of the three microphones is shifted in frequency by an amount proportional to the velocity of the emitter on the body member along the respective axis of each microphone,

$$f_d = \frac{V}{V - v} f_0$$

where f_d is the Doppler frequency, V is the velocity of sound, f_0 is the transmitter frequency, and v is the velocity of the source. The signals are fed to limiter-amplifiers and discriminators in the main console. The limiter-amplifier consists of a stage of amplification and three stages of Zener-diode clipper circuits. The output of the limiter circuit remains constant at 32 volts rms as the nominal 20-kc signal varies in level from 50 mv to 3 volts, producing a dynamic range of about 35 db. Since the signal-to-noise ratio is usually poorer than 35 db, improvement in dynamic range is of no benefit. Each signal then passes to a tuned inductance-capacitance discriminator that has a linear bandwidth for ± 300 cps. The output of the discriminator is a bipolar d-c voltage proportional to the Doppler frequency shift, which is proportional to the body-member velocity. The signals are recorded on a direct-writing oscillographic recorder. The output is also integrated and differentiated, and the resulting signals, which are proportional to body-member displacement and acceleration, are recorded. Output signals from the three limiter-amplifiers are also directly recorded on magnetic tape for storage or further processing by computers [5].

Verifying the measurements of UNOPAR was difficult because no criteria measurements were available. A one-dimensional motion was set up and measured in three dimensions by three channels. A test was designed in which the transducer was driven by a linkage in simple harmonic motion. The maximum velocity corresponding to the peak-deviation frequency was 6 ft/sec, causing a Doppler shift of about 102 cps. The modulating rate, the frequency of the simple harmonic motion, was 3 cps. The linkage driving the transducer was also connected to a magnetic plunger in a long coil. The resultant velocity from the three channels coincided with the velocity obtained from the coil to within 2 per cent accuracy, with a standard deviation of 0.6 per cent [5].

As an illustration of the measurements yielded by the UNOPAR device, consider the following experiment performed with a one-dimensional motion: The operator moved a frame on rigid rods from one side to the other for a distance of 19 in. At each end a force lever had to

be pushed aside with a pressure of 3 ± ¼ lb. Figure 37-3 shows a tape record of a typical run for a subject's first trial. Note that the breaks in the velocity curve indicate the points at which contact was made with the force lever. Figures 37-4 and 37-5 show the velocity curves for later runs for the same subject in the same experiment.

Although the three velocity curves (curves *B* in Figures 37-3 to 37-5) appear somewhat alike, it should be noted that the time has decreased

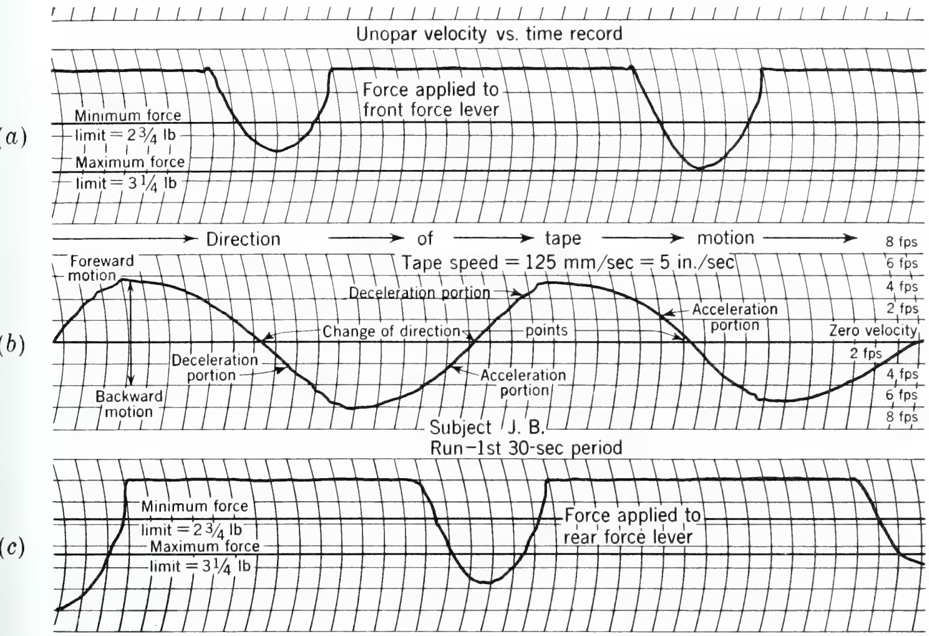


FIG. 37-3 *Tape record of first 30-sec run period of one-dimensional learning experiment. (After Gerald Nadler and Jay Goldman, The Unopar, J. Ind. Eng., vol. 9, no. 1, pp. 58–65, January–February, 1958. By permission of the publishers.)*

and a compensating increase in velocity has taken place between the first run (Figure 37-3) and the thirtieth run (Figure 37-4). The eighty-eighth run (Figure 37-5) shows that the shape of the velocity curve is about the same but that the peak velocity has decreased slightly from that of the thirtieth run (Figure 37-4) and the pattern of the velocity curve has become more rounded at the top. Other qualitative conclusions can be summarized as follows:

1. There was no change-in-direction delay. The velocity plots move directly through zero (across the center line of the velocity curve) with no pause due to the change in direction. This finding agrees in many respects with simple harmonic theory. On the contrary, classical motion-

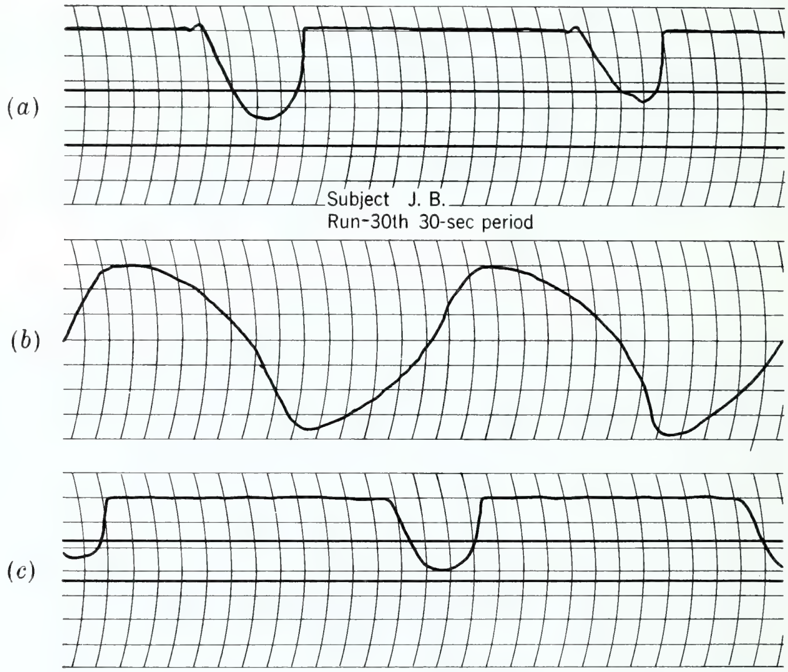


FIG. 37-4 *Tape record of thirtieth 30-sec run period.*

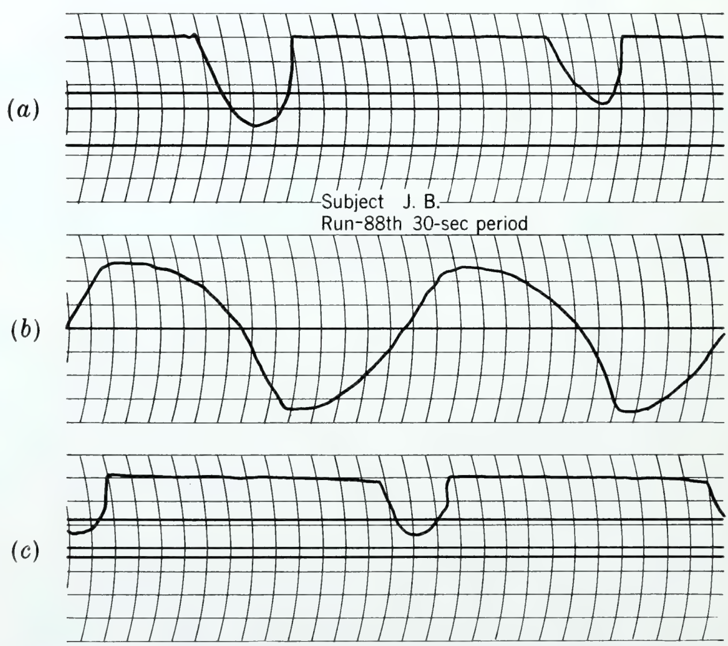


FIG. 37-5 *Tape record of eighty-eighth 30-sec run period.*

and time-study concepts indicate that delay time should be expected with a change in direction.

2. Acceleration was much slower than deceleration. (This conclusion is limited to the experimental motion of this test; other motions may provide a different pattern.) The peak acceleration in almost every cycle was much less than the peak deceleration. This is contrary to the hypothesis that most engineers accept: that a person tends to accelerate relatively quickly and then to drift or to decelerate more slowly back to zero velocity when he attempts to stop for the purpose of overcoming a force at the end of the motion.

3. There was a rapid deceleration near the point of initial contact with the force lever. In every case there was a rapid deceleration indicated by the steeply sloped, small deceleration portion of the velocity curve (the slight bump on the back side of curves *B* in Figures 37-3 to 37-5). This indicates that there was a considerable change in velocity slightly after the point of initial contact with the force lever. The velocity of the subject's hand was decreased either because the force that he was trying to overcome increased or because he recognized that he now had to start greater control of his motion.

4. There was some indication that curve shape and appearance change with increased learning. For the beginning of the operation, the velocity curves have a somewhat flat-topped appearance, with a nonsymmetrical distribution about the peak velocity point as axis. However, as the number of cycles increased, the velocity curves became more sinusoidal in shape, although they were still nonsymmetrical. There is not the flat-topped appearance, and there is an indication that the curves tended to become more symmetrical. The patterns themselves became much more smooth and rounded in their formulation, and this could be an indication of learning.

5. There was no observable constant velocity. Contrary to the classical belief that motions are comprised of a certain period of acceleration (40 per cent of the time), a period of constant velocity (20 per cent of the time), and a period of deceleration (40 per cent of the time) [1 to 3, 6], the velocity plots indicate that there is no such thing as a period of constant velocity. There is only a period of increasing acceleration up to a peak velocity point and a period of decreasing deceleration back to the zero or change-of-direction point. This finding has important implications for developments in time-study rating or pace comparison [7].

6. The above-mentioned conclusions seem to be applicable at all stages of learning. This indicates that there may be a basic motion characteristic for the activity that remains unchanged regardless of the curve-characteristic indicators that are superimposed at early stages of

learning. Information about this basic motion characteristic might provide a means for reducing the learning time on a job.

In addition to providing qualitative information, the tape records were numerically analyzed. For this analysis, the following basic and derived data were used:

1. Run number, run = 30-sec test
2. Cycles/run, cycle = one motion forward + one motion backward
3. Forward (f) force (number per run of attempts in which the force lever was pushed aside within the specified limit of $3 \pm \frac{1}{4}$ lb for forward motion)
4. Backward (b) force (number per run of attempts in which the force lever was pushed aside within the specified limit of $3 \pm \frac{1}{4}$ lb for backward motion)
5. Peak velocity for forward
6. Peak velocity for backward
7. Total forward-motion time
8. Total backward-motion time
9. Time spent in acceleration forward
10. Time spent in deceleration forward
11. Time spent in acceleration backward
12. Time spent in deceleration backward
13. Peak acceleration forward
14. Peak deceleration forward
15. Peak acceleration backward
16. Peak deceleration backward
17. Percentage of time spent in acceleration in forward motion
18. Percentage of time spent in deceleration in forward motion
19. Percentage of time spent in acceleration in backward motion
20. Percentage of time spent in deceleration in backward motion
21. Acceleration time–deceleration time for forward
22. Acceleration time–deceleration time for backward
23. All times in acceleration ($f + b$)
24. All times in deceleration ($f + b$)
25. All peak accelerations ($f + b$)
26. All peak decelerations ($f + b$)
27. All per cent acceleration times ($f + b$)
28. All per cent deceleration times ($f + b$)

Some of the more interesting graphs constructed from these data are shown in Figures 37-6 to 37-10. Figure 37-6 shows peak velocity vs. run number for subject B. K.; Figure 37-7 shows the same graph for subject J. B. These curves show that there was a slight downward trend after the peak velocity had reached its maximum value. Subject B. K.

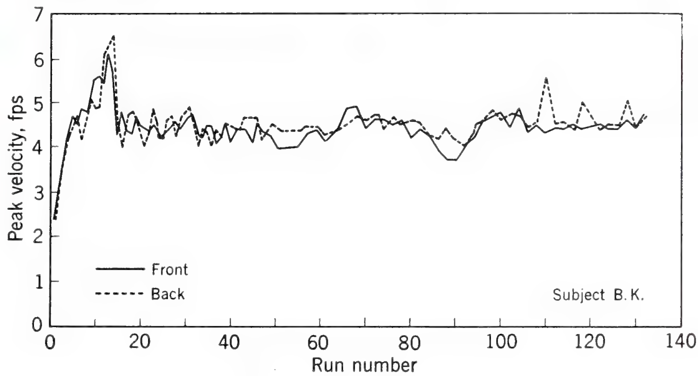


FIG. 37-6 *Peak velocity vs. run number for B. K. (After Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-plane Motion Learning, J. Ind. Eng., vol. 9, no. 3, pp. 187–197, May–June, 1958. By permission of the publishers.)*

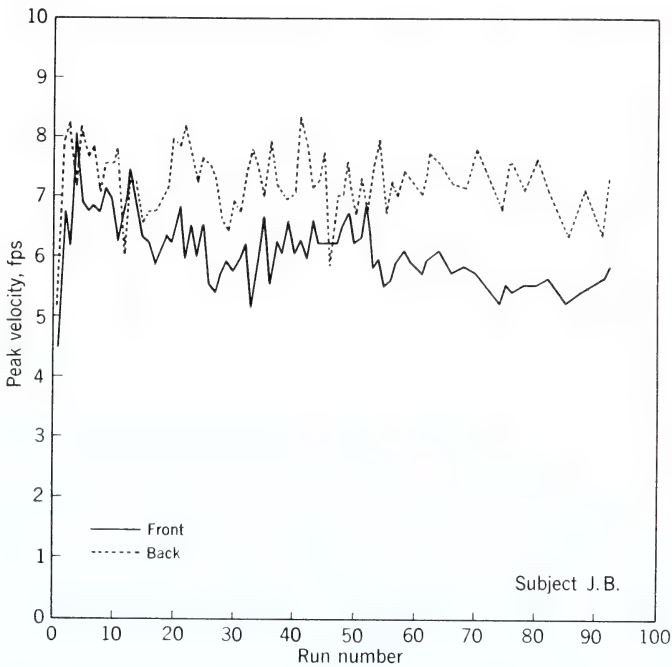


FIG. 37-7 *Peak velocity vs. run number for J. B. (After Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-plane Motion Learning, J. Ind. Eng., vol. 9, no. 3, pp. 187–197, May–June, 1958. By permission of the publishers.)*

showed a more pronounced learning maximum and drop. Also, the front-motion peak velocity decreased at a more rapid rate than did the back peak velocity. The forward motion was that of the arm (right arm) moving toward the left side of the body (the front of the body was

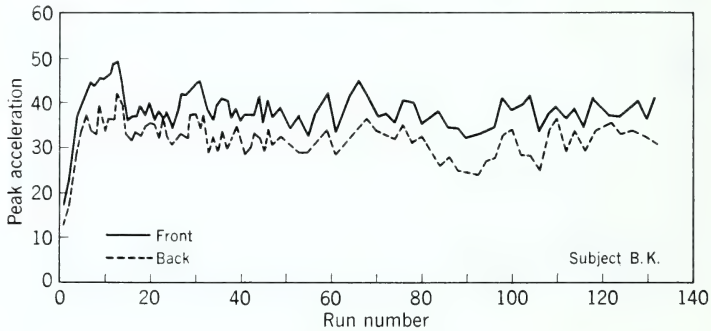


FIG. 37-8 *Peak acceleration vs. run number for B. K. (After Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-plane Motion Learning, J. Ind. Eng., vol. 9, no. 3, pp. 187–197, May–June, 1958. By permission of the publishers.)*

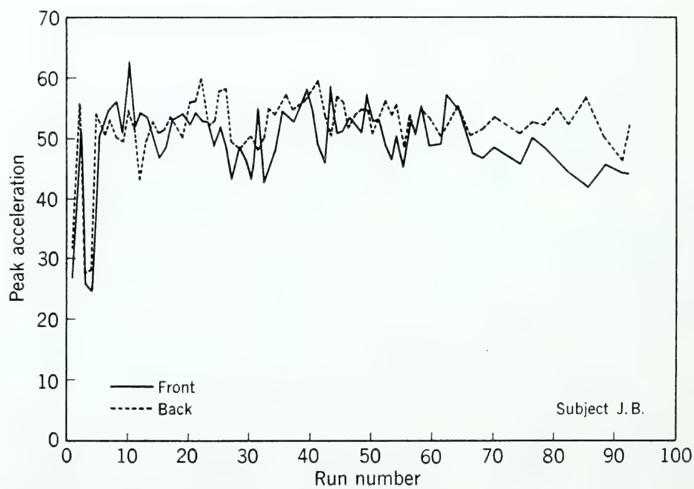


FIG. 37-9 *Peak acceleration vs. run number for J. B. (After Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-plane Motion Learning, J. Ind. Eng., vol. 9, no. 3, pp. 187–197, May–June, 1958. By permission of the publishers.)*

parallel to the vertical plane that contained the one-plane motion). This motion to the left is usually considered as the one capable of imparting the greatest force and/or velocity.

Figure 37-8 shows peak acceleration vs. run number for subject B. K.; Figure 37-9 shows the same graph for subject J. B. These graphs tend

to show the usual learning-curve effect. There was virtually no change after the first five runs. In a sense, learning was accomplished for this task at an early point. Another interesting facet concerns the marked downward trend of Figure 37-8. Also, there seems to be a cyclic behavior associated with the acceleration as the run number increases. This might indicate the presence of a cyclic phenomenon in the performance pattern of the individual.

Figure 37-10 shows attempts in which the force lever was pushed aside within the specified limit of $3 \pm \frac{1}{4}$ lb for backward motion vs.

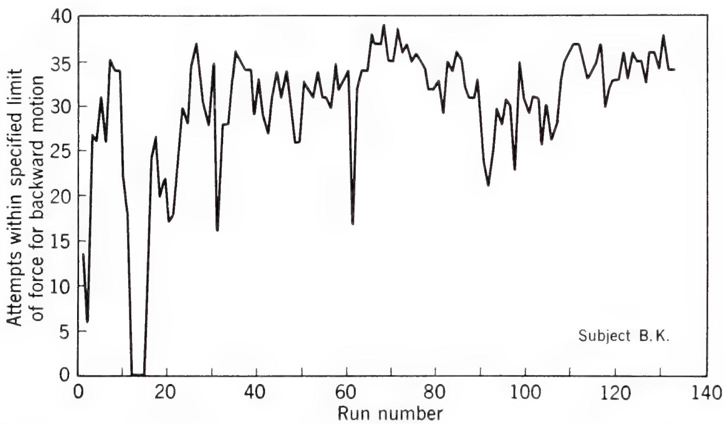


FIG. 37-10 *Attempts in which the force lever was pushed aside within the specified limit of $3 \pm \frac{1}{4}$ lb for backward motion vs. run number for B. K. (After Gerald Nadler and Jay Goldman, Operator Performance Studies: I—One-plane Motion Learning, J. Ind. Eng., vol. 9, no. 3, pp. 187–197, May–June, 1958. By permission of the publishers.)*

run number for subject B. K. The cyclic characteristic is again apparent.

In addition to these graphical analyses, Pearsonian correlation coefficients were found for pairs of the variables listed above.

For subject B. K. there was a high positive correlation between cycles per run and peak acceleration and deceleration for forward and backward motion. There was a somewhat negative correlation between cycles per run and time spent in acceleration and deceleration for forward and backward motion. (As cycles per run increase, time spent in acceleration and deceleration should decrease.) There was a significant negative correlation between cycles per run and total time for forward and backward motion. For subject J. B. there were no significant correlations between cycles per run and any of the other variables.

For subject B. K. there was a high degree of correlation between total forward-motion time and total backward-motion time. In addition,

both time values were highly correlated with time spent in acceleration forward and time spent in acceleration backward; they were not so highly correlated with time spent in deceleration forward and time spent in deceleration backward. Total forward-motion time and total backward-motion time were significantly negatively correlated with peak acceleration and deceleration for forward and backward motion, with peak deceleration forward and backward slightly higher. For subject J. B. there was a significant correlation between total forward-motion time and total backward-motion time. There was also a significant correlation between these time values and peak acceleration and deceleration for forward and backward motion, but although peak deceleration was higher than peak acceleration, it was not as much higher as it was for subject B. K. Between the total motion times and the other variables of interest here, time spent in acceleration and deceleration forward and backward, percentage of time spent in acceleration and deceleration forward and backward, and peak velocity for forward and backward motions, subject J. B. showed the same relations as subject B. K., except that some of J. B.'s correlations were not so high as B. K.'s. One striking difference between subjects J. B. and B. K. was J. B.'s negative correlation between total forward-motion time and peak velocity backward.

In general, both subjects showed the following results: Time spent in acceleration backward is correlated with peak deceleration forward. Peak deceleration forward is correlated with peak acceleration backward. Likewise, peak acceleration forward is correlated with peak deceleration backward. These results indicate that there is some relation among motions. For all practical purposes, the percentage of time in acceleration and deceleration is not related to other factors.

Some general conclusions may be summarized as follows:

1. From various graphical techniques and experimental data, it is apparent that the operators learned the job within an extremely short period of time. Although this conclusion may have been drawn from ordinary "production" data on such a job, every variable connected with the operation showed the same trend.

2. From both quantitative (correlations between cycles per run and percentage of time spent in acceleration and deceleration forward and backward) and graphical data, it appears that there is a constant percentage of acceleration time and a constant percentage of deceleration time within a motion. This is true regardless of the stage of learning. In these experiments, it appears that 38 per cent of the time was spent in deceleration and 62 per cent in acceleration (subject J. B. had 62 per cent for acceleration and B. K. 60.5 per cent).

3. There was much evidence to show that motions are interdependent and that one aspect of one motion is highly dependent on

another aspect of another motion. For example, forward force is highly correlated with peak acceleration and deceleration backward. Which is the cause and which is the effect is not known.

4. There was some indication of a downward trend in the learning process after the peak had been reached.

5. A cyclic characteristic seems to be present in some learning curves, e.g., Figures 37-8 and 37-10. This is somewhat unusual, and it may be an important characteristic of learning.

6. People are not able to perform identical motions in opposite directions with the same degree of facility. For example, backward motion was the most difficult to perform properly. The force was hard to keep within limits. This is shown by visual inspection of the curves and by the correlation between backward force and run number. Likewise, Figure 37-7 shows an interesting contradiction. The front peak velocity decreases more rapidly than the back peak velocity after reaching the plateau peak. An attempt was made to find the relation between attempts in which the force lever was pushed aside within the specified limit of $3 \pm \frac{1}{4}$ lb for forward motion and attempts within the specified limit of $3 \pm \frac{1}{4}$ lb for backward motion, but mechanical difficulties in testing prevented this.

Another experiment that utilized UNOPAR measurements investigated the effect of direction, weight handled, pace, and gravity on the acceleration pattern of the right arm. Before body-member acceleration patterns could be studied, some definite measure of these patterns had to be developed. An acceleration curve was defined as a plot of acceleration (and deceleration) vs. motion time. Similarly, an acceleration pattern was defined as an acceleration curve, with acceleration expressed as a percentage of peak acceleration and time expressed as a percentage of total motion time. An acceleration pattern is therefore a standardized curve that will not reflect differences in amount of time or amount of acceleration but will reflect the distribution of total acceleration with respect to total time. A complete measure of an acceleration pattern could become extremely detailed and lengthy. However, since velocity is a direct result of acceleration, some knowledge of acceleration patterns can be gained by studying velocity curves, as shown in Figure 37-11.

For this study the following two measures were utilized to detect fundamental changes in acceleration patterns:

$$\text{Measurement } A = \frac{\text{maximum velocity}}{\text{average velocity}} = \frac{V_m}{\bar{V}}$$

Measurement A, an index of acceleration, indicates the distribution of acceleration (and deceleration) with respect to the percentage of total

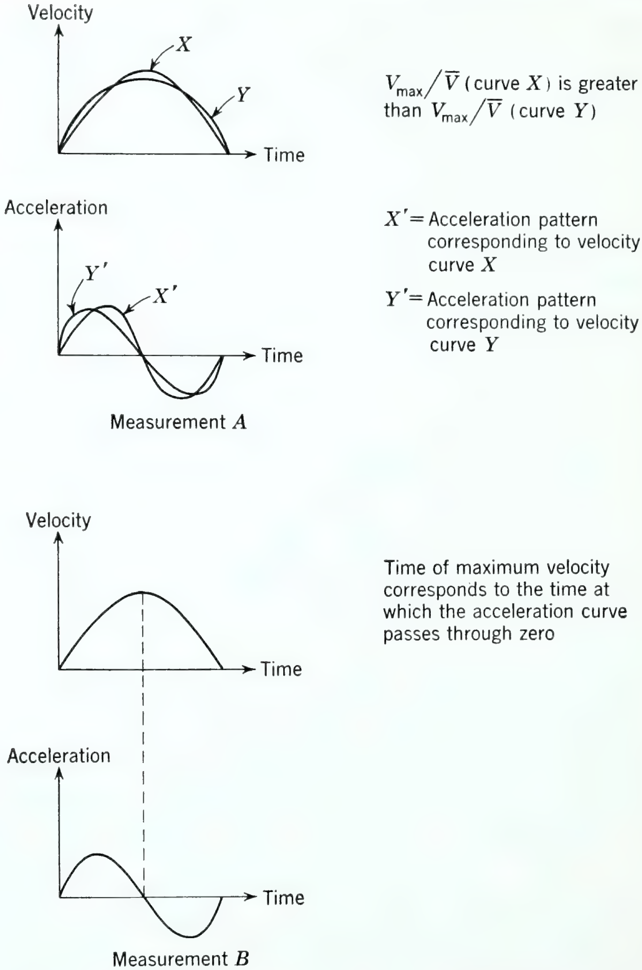


FIG. 37-11 *Acceleration-pattern measurements. (After Donald Franz and Gerald Nadler, New Measurements to Determine the Effect of Task Factors on Body Member Acceleration Patterns, J. Ind. Eng., vol. 9, no. 5, pp. 317–323, September–October, 1961. By permission of the publishers.)*

motion time. If a higher proportion of acceleration (and deceleration) occurs near the beginning (and end) of a motion, the ratio $V_m/\bar{V}$ will decrease (see Figure 37-11).

$$\text{Measurement } B = \frac{\text{time of maximum velocity}}{\text{total motion time}} = \frac{t_{v_m}}{t_t}$$

Measurement B indicates the percentage of total motion time spent in acceleration. If a higher proportion of total motion time is spent in acceleration, the ratio t_{v_m}/t_t will increase (see Figure 37-11).

A work surface containing two terminal microswitch points separated by a distance of 15 in. was used to describe a specific task. The task consisted in moving the right hand between the terminal points. Three task orientations were used. Within each task orientation, two directions were analyzed. For each direction within each orientation, two subjects each performed the given task at three levels of pace (90, 135, and 180 beats per minute) and three levels of weight (0, 14, and 28 oz). A total of 108 cells resulted from the $2 \times 2 \times 3 \times 3 \times 3$ factorial design. For each cell, the index of acceleration (measurement A) and the proportion of total motion time spent in acceleration (measurement B) were obtained after the subjects were trained.

For orientation 1 the work surface was located in a horizontal plane so that a line through the terminal points was parallel to a line through the shoulders of the subject. Within orientation 1, direction 1 consisted in a movement to the left, and direction 2 was to the right. For orientation 2 the work surface was located in a horizontal plane so that a line through the terminal points was perpendicular to the vertical plane containing the shoulders of the subject. Within orientation 2, direction 1 consisted in the movement inward or toward the body, and direction 2 was the movement outward. For orientation 3 terminal points were located on a line perpendicular to a horizontal plane. The plane of the work surface was parallel to the vertical plane containing the shoulders of the subject. Within orientation 3, direction 1 consisted in the movement downward, and direction 2 consisted in the movement upward.

The average ratio of maximum velocity to average velocity for the entire experiment was 1.47 (index of acceleration). The average value of per cent time spent in acceleration was 51.8 per cent (proportion of total motion time spent in acceleration). Other results were:

1. All three orientations were significantly different with respect to the index of acceleration. Orientation 3 (left-right movement) was the largest and orientation 2 (in-out movement) was the smallest in relation to the index of acceleration. No significant difference was found in the proportion of total motion time spent in acceleration as a result of orientation.

2. Within orientation 3 (up-down movement), movement upward was significantly smaller than movement downward in relation to the index of acceleration. Within orientation 3, movement upward was significantly greater than movement downward in the proportion of total motion time spent in acceleration.

3. There was a significant difference between the index of acceleration and the proportion of total motion time spent in acceleration. Subject 2 had the smallest index of acceleration and the largest proportion of total motion time spent in acceleration. Subject and orientation interaction was highly significant.

4. Pace had a significant effect on the index of acceleration and the proportion of total motion time spent in acceleration. With increase in pace, the index of acceleration decreased linearly and the proportion of total motion time spent in acceleration increased linearly. The interaction between pace and direction within orientation was significant in the proportion of total motion time spent in acceleration. Although some significance was found in all orientations, the most definite significance was in up-down movement. In all orientations and directions there was a trend for the proportion of total motion time spent in acceleration to increase with increased pace.

5. Weight had a significant effect on the index of acceleration. With increase in weight, the index of acceleration decreased linearly. The interaction between weight and direction within orientation was significant in the proportion of total motion time spent in acceleration. This significance was located only in orientation 3 (up-down movement). In orientation 3 an increase in weight linearly decreased the proportion of total motion time spent in acceleration direction and tended to increase it in the upward direction.

6. Although not statistically significant, an interesting trend was found in the pace-weight interaction. A given increase in weight caused an increasing reduction in the index of acceleration as pace increased. A given increase in pace caused a decreasing reduction in the index of acceleration as weight increased.

The following limitations apply to the conclusions obtained in this experiment:

1. All data in this experiment were obtained from two male subjects.
2. The subjects used were volunteers from the senior industrial-engineering class, and no attempt was made to obtain "average" subjects.
3. The experiment consisted of a right-hand activity only.
4. The basic task studied consisted in a simple motion containing a minimum of inherent difficulty factors.
5. Pace was limited to three specific levels, 90, 135, and 180 beats per minute. Weight handled was limited to 0, 14, and 28 oz.
6. All arm motions were measured at a point directly on top of the subject's wrist. This required the use of a different distance factor for each orientation.

The following were concluded within the above limitations:

1. Acceleration patterns of different people are affected in the same general manner by pace, weight, direction, and gravity.
2. Acceleration patterns can be expected to differ significantly between individuals because of individual differences.
3. Weight handled and direction of motion (relative to the gravitational force) significantly affect the index of acceleration of the body-

member acceleration pattern. Pace also has a highly significant effect on the index of acceleration. Increases in pace or weight decrease the index of acceleration. An increase in the amount of gravitational force in opposition to body-member acceleration also causes a decrease in the index of acceleration.

4. Pace has a highly significant effect on the proportion of total motion time spent in acceleration. Increased pace results in increased proportion of total motion time spent in acceleration, regardless of orientation or direction.

5. Direction (relative to the gravitational force) has a highly significant effect on the proportion of total motion time spent in acceleration. The proportion of total motion time spent in acceleration is less when acceleration and gravitational force are in opposition. When acceleration is aided by the gravitational force, the proportion of total motion time spent in acceleration is greater.

6. The interaction of weight and direction (relative to the gravitational force) has a highly significant effect on the proportion of total motion time spent in acceleration. However, weight alone has no significant effect on this measure. When gravity aids acceleration, increased weight decreases the proportion of total motion time spent in acceleration. When gravity and acceleration are in opposition, increased weight increases the proportion of total motion time spent in acceleration.

These conclusions lead to some interesting observations and conjectures. For example, the result showing that a body member chooses different acceleration patterns for different levels of pace and weight and for different directions (relative to the force of gravity) is logical, for, if for a given task the pace is increased to a maximum, a greater amount of acceleration must be put forth earlier in the motion (when the speed of muscular contraction is low) in order to accomplish the given task. Although past research is capable of explaining why some change in acceleration pattern must occur, not enough evidence is available to provide an explanation of why one acceleration pattern is chosen in preference to others that are available to the body member within the limitations of external force available. A possible answer to this question is that, to perform a given task, the body seeks a motion pattern that is optimum in the light of all factors affecting the body and that, with training, the body member closely approaches this optimum.

BIBLIOGRAPHY

1. Barnes, R. M.: "Motion and Time Study," 4th ed., pp. 238–239, John Wiley & Sons, Inc., New York, 1958.

2. ——— and M. E. Mundel: Studies of Hand Motions and Rhythm Appearing in Factory Work, *Univ. Iowa Studies Eng. Bull.* 12, 1938.

3. Deegan, W.: The Development and Use of a New Technique for Measurement in Time and Motion Study, M.S. thesis, University of Iowa, Iowa City, Iowa, 1935.

4. Goldman, J.: Development and Testing of an Electronic Method for Determining Acceleration, Constant Velocity, and Deceleration of Body Motions, D.Sc. dissertation, Washington University, St. Louis, Mo., August, 1955.

5. ——— and D. K. Ross: Measuring Human Work Performance, *Electronics*, vol. 34, no. 10, pp. 196–198, Mar. 10, 1961.

6. Mundel, M. E.: "Motion and Time Study," 2d ed., p. 379, Prentice-Hall, Inc., Englewood Cliffs, N.J., 1955.

7. Nadler, G.: "Motion and Time Study," Chap. 23, McGraw-Hill Book Company, Inc., New York, 1955.

8. ——— and J. Goldman: Apparatus for Determining Velocity of Human Body Members, U.S. patent 2,914,730, Nov. 24, 1959.

Chapter 38

DOCUMENTATION OF HUMAN FACTORS ENGINEERING DATA

Paul G. Ronco, Donald B. Devoe,† and Ezra V. Saul**

ONE OF THE MOST PRESSING NEEDS OF THE MODERN SCIENTIFIC COMMUNITY is that of efficient communication of scientific theories and research data. Awareness of existing information is essential, not only for preventing needless duplication of effort, but also for extending the benefits that may be derived from knowledge of the experience of many researchers. In this era of rapid scientific advances the need for adequate scientific communication is becoming increasingly critical and more difficult to satisfy. No longer is it possible for the individual investigator to keep up with all the technical information in a given field.

INFORMATIONAL NEEDS OF THE HUMAN FACTORS SPECIALIST

The type of subject matter encountered in the literature of human factors engineering is determined by the several needs of the human factors specialist. First, he needs a broad range of information to aid him in specific problems of systems design and development. He requires information about factors that influence the efficiency of man-machine interaction: personnel-selection techniques, training methods, operator-oriented design of equipment, and the social conditions of the work situation. Solution of man-machine problems requires extrapolation of data from such areas as experimental, social, educational, and industrial psychology, anthropometry, biology, physics, medicine, and the many areas of engineering. It is notable that often it is the research of a

* Institute for Psychological Research, Tufts University, Medford, Mass.

† Sperry Rand Research Center, Sudbury, Mass.

This chapter was prepared in part under the joint sponsorship of the Engineering Psychology Branch, Psychological Sciences Division, Office of Naval Research, Contract Nonr 494(13), the Office of Scientific Research, U.S. Air Force, and the Office, Chief of Research and Development, Department of the Army.

scientist who is himself not interested in human factors engineering that makes an important contribution to this area. Second, the human factors specialist requires information on research methodology as well as empirical data. This need for methodology is especially great when specific experimental data are not available and the specialist plans to conduct the research himself to obtain the necessary data.

The information required by the human factors specialist is published on several levels of technical complexity, ranging from the popular magazine article to the technical handbook. An additional difficulty is introduced by the peculiarity of the language used in any professional field. Because of these communication problems, the human factors specialist has a third major need: a form of dictionary service that enables him to translate the information provided by non-human-factors research into the idiom of human factors literature and to express his own problems and findings in these terms. Furthermore, the human factors specialist often needs consultants to explain and evaluate data that are unfamiliar to him. This need is apparent particularly in dealing with the literature of psychology, which is heavily infused with terminology unknown to the nonpsychologist. As observed by the Board of Scientific Affairs of the American Psychological Association [1], the methods of technical communication in psychology have evolved from the experience, conditions, and needs of the past, without much deliberate planning or self-evaluation. However, there is little doubt that engineering concepts are currently exerting a marked effect upon psychological terminology and that the future will provide uniformity in the language used in human factors literature and coordination of concepts among the various technical fields involved. It is to be hoped that what will evolve is a literature that reflects the needs of the user of the information rather than those of the originator.

STATE OF THE ART IN HUMAN FACTORS DOCUMENTATION

The human factors literature is in its present state extremely useful. There are a great number of data in the area of sensory processes, motor outputs, complex thought processes, selection, and training. There is an increasing amount of quantitative information about the human being's capacity to perform under unusual environmental conditions, such as acceleration forces, extreme temperature and humidity, extreme illumination and noise, and sensory deprivation. An extensive literature has been developed pertaining to the human factors design and evaluation of specific equipment and systems, the findings of which often have explicit transfer value in the development of new prototype equipment. Furthermore, literature exists concerning theoretical models and general

equations for describing and predicting operator-equipment interactions that is proving useful to those responsible for systems design and development. Also there are many handbooks and general publications, documents that give comprehensive coverage of broad areas and/or present information applicable to the "quick-fix" solution of problems. A bibliography of such literature is given in Appendix I. It must be emphasized, however, that there are no entirely adequate handbooks of human factors information that are directly usable by the design engineer.

The informational problems in the human factors area have certainly not arisen from any dearth in the amount of literature generated. It is probably safe to say that there are at least 5,000 documents produced a year that have direct relevance to this area. However, the prodigious growth in the number of human factors publications has created a serious problem. The utility of the information is, of course, a function of its availability to the human factors specialist. The problem is: Where does one go to find this literature?

There are three dichotomies that characterize all human factors publications. One of these is the division of military vs. nonmilitary literature. Probably as much as 70 per cent of the English-language human factors documents are in the form of military or military-sponsored reports, published by government laboratories or government-supported university and commercial organizations. The second dichotomy is that of security-classified vs. unclassified material. Fortunately only a small fraction of the total human factors literature bears any security classification, and of the classified material very little appears at higher classification levels than Confidential. The third dichotomy by which the sources of human factors data may be classified is that of foreign vs. domestic publication.

There are already at least 300 English-language journals publishing human factors material, and there are probably just as many foreign journals, in many different languages, that also present such information. Of particular interest at this time is, of course, the Russian literature. A 1959 survey of available Russian periodicals revealed that there are at least 15 Soviet journals that produce a large amount of human factors material [12]. Some of the more important English-language journals that have been found to yield much information relevant to human factors engineering are listed in Table 38-1.

Access to the titles of nonmilitary articles and books on human factors is achieved primarily through a search of such bibliographic sources as *Applied Science and Technology Index*,[1] *Biological Abstracts*,[1] *Business Periodical Index*,[1] *Current Index to the Medical Literature*[1]

[1] Addresses of these journals may be found in the latest edition of "Ulrich's Periodicals Directory," published by R. R. Bowker Company, New York.

TABLE 38-1 *Selected English-Language Journals Producing Human Factors Information**

Aerospace Medicine
American Journal of Psychology
Annals of Occupational Hygiene
Aviation Week
Engineering and Industrial Psychology
Ergonomics
Human Factors
IRE Transactions on Human Factors in Electronics
Journal of the Acoustical Society of America
Journal of Applied Physiology
Journal of Applied Psychology
Journal of Experimental Psychology
Journal of the Optical Society of America
Occupational Psychology
Operations Research
Perceptual and Motor Skills

* Mailing addresses of these journals may be found in the latest edition of "Ulrich's Periodicals Directory," published by R. R. Bowker Company, New York.

(formerly *Index Medicus*), *Psychological Abstracts*,[1] and *Science Abstracts*.[1] The titles of documents published under the government aegis may be obtained in the *U.S. Government Research Reports*,[2] the Armed Services Technical Information Agency's (ASTIA) *Technical Abstract Bulletin*,[3] the *Technical Reports Newsletter*,[4] and the *AEC Reports Price List*.[4] Access to titles of articles appearing in foreign journals is a little more difficult. Especially helpful guides to foreign literature available in this country are "A Guide to U.S. Indexing and Abstracting Services in Science and Technology" [3] and "Translators and Translation, Services and Sources" [8]. Lists put out by the Office of Technical Services, Department of Commerce, give titles and abstracts of translated technical literature available from the Office of Technical Services, Library of Congress, the Special Libraries Association, cooperating foreign governments, commercial translators and publishers, universities, and several other sources. Some major abstracting periodicals, such as *Biological Abstracts*, review foreign literature; the *Scientific Information Report*[5] is also useful as a bibliography of foreign publications. The *Monthly Index of Russian*

[2] The titles of unclassified, unrestricted documents listed in the ASTIA *TAB* are made available to the general public in the *U.S. Government Research Reports*.

[3] Obtainable by subscription from the U.S. Government Printing Office.

[4] Obtainable from the Office of Technical Services, Department of Commerce, Washington.

[5] Obtainable from the Office of Technical Services, Department of Commerce, Washington.

Accessions[5] is the major U.S. bibliography of Soviet publications; the *Bibliography of Agriculture*[6] and *Current Index to the Medical Literature* also include Soviet material. Along with the increasing interest in foreign scientific literature are developing many commercial establishments for translating, abstracting, and indexing the foreign literature in many fields.

Obtaining articles or documents is, of course, much more difficult than finding their titles. Access to much of the material handled by ASTIA, i.e., research sponsored by the Department of Defense, depends upon whether the requester of the information is himself working on a government contract, whether he has a "need to know," and often whether he has a security clearance. Access to nonmilitary literature is determined primarily by the availability of well-funded and competently staffed technical libraries. Unfortunately, even in the libraries of such well-endowed institutions as Harvard University and the Massachusetts Institute of Technology there is found only a small fraction of the journals and books pertinent to human factors engineering. Furthermore, the availability of human factors literature, both military and nonmilitary, is sometimes restricted by industrial laboratories themselves, through such classifications as "Proprietary Interest" and "Company Confidential."

ORGANIZATION OF HUMAN FACTORS DATA-INDEXING SYSTEMS

The ordering of available data is becoming more and more of a problem for the human factors specialist concerned with organizing the information in the field. First, the data organizer must acquire as much of the literature as possible from journals, government laboratories, universities, and industrial sources. Second, there is the task of translating into English the large bulk of material from foreign literature. Third, and most important, the organization of the data requires the elaboration of an easily understood and useful classification scheme and the coding of the document acquisitions into that scheme.

There are several criteria by which an indexing system should be evaluated. First of all, an indexing system must provide rapid, accurate, and complete retrieval of desired material. It must be an efficient scheme for the manual location of literature and should be adaptable to machine classification of information. The system must take into account the heterogeneity of the literature and the dissimilarity among users in terms of training, experience, and orientation. Furthermore, in such a constantly changing area as human factors engineering the data-classification system has to be flexible enough to adapt to the expansion of the field;

[6] Available at the Library of Congress, Washington.

the scheme must be evaluated in terms of the ease of future revision. An index should be sufficiently cross-referenced to enable the user to extend the scope and thoroughness of his search. In this way not only will additional data be provided, but also the concept being searched will be clarified. Finally, the index should present a comprehensive picture of the whole field, its scope, its components, and its interrelations.

On the basis of these criteria, it is possible to evaluate various indexing systems that may be employed with human factors data. For example, let us briefly compare three types of general classificatory systems, alphabetical indexing, classification indexing, and coordinate indexing.

Alphabetical indexing of subject headings provides the most rapid method for locating a defined subject. For example, an alphabetical index of search terms is well suited to meet the aforementioned need for a dictionary service. However, this type of indexing provides no grouping of related data to aid in a search for complete information about a subject area. Therefore, alphabetical arrangement is useful primarily as a means of ordering major headings and subheadings in other, more informative systems of classification.

Classification indexing, the arrangement of headings in logically grouped categories, permits the location of related ideas. It represents the various aspects of the indexed field and may frequently provide ideas and leads to searchers who are unfamiliar with the subject area. On the other hand, the freedom to associate concepts is often restricted in classification indexing, for classification indexing requires predetermination of which concepts will be associated and in what order they will be grouped; the system might well fail when an unanticipated combination is required. For example, the combination KNOBS, USE WITH DIALS may have been anticipated and indexed, whereas the required combination, KNOBS, USE IN EXTREME COLD, may not have been. Therefore, both KNOBS with its various pertinent subheadings and COLD with its subheadings must be searched and the resulting data screened for information pertaining to the use of knobs in extreme cold. Furthermore, although it may make a generous allowance for expansion of the indexed field, classification indexing is characterized by the limited freedom with which new ideas may be incorporated. Sometimes provision is made for the addition of low-order subcategories, but essentially the framework of major concepts is fairly rigid.

A third type of indexing system, coordinate indexing, allows freedom to associate concepts in any combination desired for the problem at hand. The Uniterm system is an example of coordinate indexing. This kind of system has the advantage of requiring no association of concepts prior to coding material. Thus, if all material related to dials is indexed under DIALS, if the material pertinent to knobs is indexed under KNOBS,

and if material related to cold is indexed under COLD, it is possible by comparing the material in these three categories and noting common references to locate literature having to do with knob-dial relations or with the use of knobs or dials in extreme cold without having had to anticipate these specific associations. Coordinate indexing offers no predetermined associations that may provide the searcher with clues for retrieving desired material or extending the scope of his search. However, this deficiency is partially compensated for by the system's adaptability to unanticipated search areas and to expansion of the indexed field.

In the development of any indexing system the advantages and disadvantages of the three methods discussed here should be considered. The choice of a literature-classification system also depends upon such factors as the nature and amount of literature to be covered, the maturity of the area in which the literature appears, and the needs of the intended users. A satisfactory system for a particular user would generally involve elements of two or even all three of the types of indexing systems that have been described. One further point should be stressed. No matter what characteristics are designed into the index, they cannot compensate for superficial, uninformed, or incompetent coding of the documents into the index.

ASTIA indexes some of the human factors literature; however, because this is only one of the many subject areas under its scrutiny, coverage is necessarily scant. A number of other bibliographic classification systems have been developed by organizations concerned exclusively with human factors engineering. One example of such a system is a punched-card method that has been described in an article in *Human Factors* [2], which presents information that can well serve as a guide to anyone interested in establishing a human factors library. Usually such organizational indexing systems are devised to meet the demands of a small, specialized group.

TUFTS UNIVERSITY HUMAN ENGINEERING INFORMATION AND ANALYSIS SERVICE

There is one indexing operation and document repository devoted solely to providing bibliographic information in the area of human factors and geared toward serving the general human factors population, Tufts University's Human Engineering Information and Analysis Service, sponsored jointly by the three military services. This project was initiated in order to overcome the problems of acquiring and integrating the widely scattered literature. It was devised with the expectation of

providing services to human factors specialists in general rather than to a smaller, more specialized group such as might be found in a single industrial concern.

The Tufts indexing system centers about a topical outline (reproduced in abridged form in Appendix II). The outline has 15 major headings, with secondary and tertiary headings supplementing the major areas. Although these headings are primarily applied in character, they are adaptable to basic literature as well. The outline consists of a set of logically related descriptors that reflects the structure of the field of human factors and permits a systematic, detailed search of the literature.

The indexing system utilizes manual retrieval, and the topical outline is used as a basis for providing detailed, numbered category headings for a subject-matter card file. After each document has been coded to all relevant headings, its citation is microfilmed together with an abstract of the document. The filmstrip is mounted on an aperture card, each filmstrip containing 10 citations and their abstracts, and the cards are placed in the file behind the relevant headings. Thus ready-made bibliographies may be retrieved by simply going to the desired headings and removing the cards filed there.

Ordinarily flexibility is somewhat reduced when an indexing system is built around a topical index such as this one. However, the Tufts index categories are not tied to a predetermined, formal numerical system, and therefore headings may be inserted, deleted, or shifted at will, provided only that the topical outline is modified to show these changes. Periodic revision of the outline has thus far given adequate flexibility to the system.

Since each document is identified and filed by a single accession number, filing and retrieval can be accomplished by unskilled clerks and the numbering permits ready conversion to machine-retrieval systems. With accession-number shelving, browsing through the documents is usually unrewarding, but the use of author and source card files in conjunction with the subject-matter file permits highly productive browsing through the cards—a procedure that also has the advantage of keeping searchers away from the document shelves, thus reducing the possibility of misfiled and lost documents.

The topical outline and the subject-matter file alone do not provide all the retrieval aid necessary in an index of human factors literature. Although the subject headings give information about problem areas, they do not adequately facilitate tracking down literature relevant to a highly specific problem. Therefore the Tufts system contains an alphabetical index of specific search terms coded to the topical-outline categories (see Table 38-2). The purpose of the alphabetical index is to

TABLE 38-2 *Sample of Tufts Alphabetical Index of Human Factors Engineering Search Terms*

Search term	Topical-outline category
AAP Complex Coordination Test	7.6.7—Equipment and methods in HE research on basic motor performance
Abdominal—extension measures	7.2.1—Body size, stationary
Aberrations	
chromatic	3.15.1—Individual differences and anomalies
optical	3.15.1—Individual differences and anomalies
Absolute pitch	4.9.1—Basic attributes: pitch
Absolute thresholds (see Thresholds)	
Acceleration	
as coding stimulus	5.5.2—Coding and signalling through kinesthesis
	5.6.1—Basic processes and data
effects on performance	12.4.1—Speed and acceleration
measurement of	12.4.0—Motion
	12.9.0—Special equipment and methods utilized in the study of the effects of special environments on performance
Acceptability	
of equipment	13.3.2—Acceptability of equipment and/or task
of food	13.5.2—Diet, food, and nutrition
of task	13.3.2—Acceptability of equipment and/or task
Access dimensions	10.2.0—Workplace design
	10.2.3—Stowage

incorporate every term that experienced coders consider useful in describing problems to which the data of the document being coded might be relevant. The further correlation of these terms to each relevant document is beyond the scope of the effort at Tufts. The alphabetical index provides translation of terms by including jargon used in fields that overlap into human factors and associating all these specialized terms with the relevant terms in the topical outline. Thus a searcher may begin by looking up words descriptive of his problem in the alphabetical index, find the appropriate term in the topical outline, and then retrieve the relevant documents.

For the purpose of aiding the user in determining what literature should be read, all documents incorporated into the system are abstracted after having been examined by the professional staff of the project. No attempt is made in the abstract to summarize data or present conclusions.

Over the past 7 years the project has acquired some 20,000 docu-

ments pertinent to human factors. At present it is incorporating 2,000 to 2,500 documents per year, obviously far fewer than the estimated 5,000 or more articles being produced annually. Limitation of the number of documents indexed is a financial and manpower limitation rather than a lack of available material; the Tufts project can procure a larger proportion of the published material than it can process.

IMPLICATIONS FOR THE FUTURE

What are the implications of the present state of human factors documentation for the potential user? It is obvious that the user will have to search several sources for bibliographic information. He should be familiar with several major reference works, such as the McCollom and Chapanis bibliography [10], the annual Tufts "Human Engineering Bibliography" [4 to 7], the "Handbook of Human Engineering Data for Design Engineers" [9], the "Human Engineering Guide for Equipment Designers" [13], and "Human Engineering for Equipment Design" [11]. If the user is involved in the design of equipment for military use, he should procure copies of military specifications dealing with human factors. There are specifications presenting information about human factors in relation to guided missiles, fixed-wing aircraft, manned aircraft, air weapons, and other military systems. Other specifications deal with such human factors areas as control configurations and markings, lighting equipments, training equipments, communications equipments, cockpit layout, etc. The user should certainly have available the journals listed in Table 38-1. There are many other excellent reference works concerned with human factors, some of which have been listed in Appendix I. Also, there is an increasing number of "handbook"-type publications being produced that should be of aid to human factors specialists. Of course, the specialist can do much to advance the cause of documentation efforts in his own field by committing his human factors research to writing and transmitting his reports to major documentation centers.

Today the great cry in documentation is for mechanized information-retrieval systems. Unquestionably, bibliographic efforts in the field of human factors engineering will become mechanized. For instance, material from the 1957, 1958, and 1959 Tufts Human Engineering Bibliographies has been indexed in the Key Word in Contex (KWIC)[7] system by a computer. But, however desirable mechanized documentation may be, it alone cannot solve the major problems now faced by

[7] Personal communication, R. S. Hirsch, International Business Machines Advanced Systems Development Laboratory, San Jose, Calif.

documentation efforts in human factors engineering. The present volume of literature does not seem to warrant the investment necessary for the mechanization of documentation in this area. What is more urgently needed are efforts to increase the centralized acquisition, classification, abstracting, and dissemination of the human factors literature, especially that published in foreign languages.

BIBLIOGRAPHY

1. American Psychological Association, Board of Scientific Affairs: Technical Communication in Psychology: A Statement of the Problem, *Am. Psychologist,* vol. 14, no. 6, pp. 267–271, June, 1959.

2. Evans, Carol S.: A Marginal Punched Card System for Human Factors Literature, *Human Factors,* vol. 1, no. 4, pp. 32–46, November, 1959.

3. A Guide to U.S. Indexing and Abstracting Services in Science and Technology, *Natl. Federation Abstracting Indexing Services Rept.* 101, June, 1960.

4. Human Engineering Bibliography (1955–1956), *ONR Rept.* ACR 24, October, 1957.

5. Human Engineering Bibliography (1956–1957), *ONR Rept.* ACR 32, October, 1958.

6. Human Engineering Bibliography (1957–1958), *ONR Rept.* ACR 43, October, 1959.

7. Human Engineering Bibliography (1958–1959), *ONR Rept.* ACR 55, October, 1960.

8. Kaiser, Frances E. (ed.): "Translators and Translations, Service and Sources," Special Libraries Association, New York, 1959.

9. Kennedy, J. L., et al.: Handbook of Human Engineering Data for Design Engineers, *ONR SDC Rept.* 199-I-1, December, 1949 (1st ed.), 1951 (2d ed.).

10. McCollom, I. N., and A. Chapanis: Human Engineering Bibliography, San Diego State College Foundation, San Diego, Calif., 1956.

11. Morgan, C. T. (ed.): "Human Engineering for Equipment Design," under contract to McGraw-Hill Book Company, Inc., New York.

12. Sawyer, Herbert L.: A Survey of Russian Literature Related to Human Engineering, Institute for Applied Experimental Psychology, Tufts University, Medford, Mass., September, 1959.

13. Woodson, W. E.: "Human Engineering Guide for Equipment Designers," University of California Press, Berkeley, Calif., 1954.

Part I

CONCLUDING COMMENTS

Chapter 39

THE SYSTEM SYSTEM

J. C. R. Licklider[*]

OFTEN WE REMIND OURSELVES THAT TECHNOLOGY IS ADVANCING AT AN exponential pace. We see it harnessing atomic energy and conquering space. Reviewing only a few years, we recall the development of television, the automation of the great network of telephones, the vanquishing of time and distance through speed in the air and endurance beneath the sea. We consider SAGE, TIROS, MIDAS—the whole parade of military supersystems that constitute man's greatest work in organizing complexes of machines and brains to fulfill stated missions. We view these many things with pride in their achievement, but also with a deep concern.

As we perceive it from our present distance, the engineering of two decades ago was dominated by equations and by rules. The products of technology were clear-cut, definite structures and devices. If they were based on theory, it was straightforward physical or chemical theory—linear circuits, Carnot cycles, autocatalytic reactions. If they were rooted in empiricism, the empiricism was conservative: the fastest airplane flew only a little faster than its predecessor; the strongest steel was but a slight variant of another steel. Technology was the application of the hard sciences. Its products were buildings, bridges, motors, and machines. Engineers shared with physical scientists and mathematicians evident pride in explicitness of method, tractability of subject matter, and definiteness of result.

The major theme is now changed. The problem is no longer to design a pulley or a gear. It is to find a mission worthy of a million men, to plan a flow of metal and ideas and of flexibility and change, to build adaptive means, to forestall obsolescence till the maiden flight is flown. The methods and techniques that served so well in making gears, that fixed the connotation of the art as sure and sharp, are to the modern

* Bolt Beranek and Newman, Inc., Cambridge, Mass.

The research on which this chapter is based was sponsored largely by the Behavioral Sciences Division, U.S. Air Force Office of Scientific Research, Office of Aerospace Research, through Contract AF 49(638)-355. Except for minor editorial changes, this chapter is a reproduction of *USAF Office of Sci. Research Rept.* 1673.

problems less relevant. Modern technology is as much politics and sociology as physics and chemistry. It is even more economics and psychology. No one knows how to partition successfully the involvement of the softer sciences from the involvement of the harder ones. Thus the central problems are problems of "organized complexity." To solve such problems we need new methods and techniques. How shall we find them?

What changes must there be? The key lies in the phrase *to organize complexity*. Modern systems are made of men and models more than metal. We who with doctrine join the men and metal often do not join them well. We have in hand no way to join them better, and therefore we rarely sense our poverty of skill. The art that we need—the art that we must devise—is one to yield coherence, to relate the workings of the parts to system goals. It is an art of aims and purposes, of controlled development. The changes that we require are changes, not of grams or dynes, but of plans and correlations. The changes must first provide clear models and then make sure that the systems that we produce meet those designs. Requirements vary, year by year, and detailed plans must follow, day by day, the vagaries of new solutions over which no rigid schedule can prevail. To harmonize great projects thus demands an agent, flexible and fast as well as strong and wise.

INFORMATION NEEDS IN VARIOUS PHASES OF SYSTEM DEVELOPMENT AND OPERATION

The domain of systems still resists analysis. No simple scheme has yet compressed the art of system making to a rigid mold. Some systems rise from definite attempts to formulate good goals and missions. A major need is seen, and ways of meeting it are then contrived. Such ways are but hypotheses, for they require procedures yet to be devised and equipment not yet realized in working form. To define a mission well requires that many ways be cross-compared to find the best, that each hypothesis be brought to test—before machines are built. If this initial step were taken well, if system aims were surely right, then costly later steps would not be stray. The gain from saving wasted steps would pay for careful mission formulation.

We have techniques—mission analysis, task analysis, time-line studies, information-flow charts, logic diagrams, and all the rest. Each such technique serves but an aspect of the whole demand. Each new result must pass from brain to brain through papers transferred hand to hand. Coordinating the paper is as hard as designing the system. The overall schema (mission and system) for a major project may be the

product of six dozen men and six months' work. It is a bare schema, reaching down only a level or two into detail. That all the parts will mesh and in operation lead to a desired result is guaranteed by intuition only, not by precise calculation. The gear, wheel, straightedge, and compass play their roles as symbols, but they do not lead through logic from premise to consequence, nor do they set forth demonstrations.

The bare, untested schema now gives rise to a vast project. A thousand men, and then ten thousand, transform each block into a hundred blueprints. They bring their gains from past encounters, apply results of past research. They construct mockups, test components, simulate parts *G* through *J*. As they transform plans into realities, they create myriad new patterns in lower levels of the structure that are individually comparable in complexity with the whole of the initial plan.

In this great expansion of the picture, the developers of the system are guided by the original proposal, memoranda, change sheets, progress reports, and personal discussion all up and down the line. However, the time constant of propagation of guidance information lengthens as the project grows. It varies from perhaps 2 days at the outset (2 days between the time when p per cent of the people who should know of a decision have not yet heard of it and the time when only p/e per cent have not yet heard) to 3 months at the peak of effort. Such guidance cannot truly guide. The main sources of information at each level of the project, therefore, are of necessity either at that level or outside the project.

Reports passing upward through the hierarchy are slower and less adequate than guidance information passing down. Whereas guidance is incomplete mainly because the source of guidance does not command details, reports are incomplete for several reasons. One cannot report technical work in full in writing and also carry on the work. No higher level can absorb or use the piles of information that would be required to represent in full the work of lower levels. Since substantive technical work often leads to a tangible result, whereas progress reports (for reasons given) rarely do, the rational worker delays reports and presses on with wire and solder.

The body of research that influences the development of a large man-machine system typically includes: (1) *ad hoc* "research" conducted as part of the system development to improve or evaluate components and subsystems; (2) applied research conducted outside the particular system project but with the aim of providing improved components or subsystems, or of advancing other aspects of technology, for application in system developments; (3) basic research that happens to be relevant. A fundamental difficulty inhibits effective system applica-

tion of the last two of these forms of research. The domain of possibly applicable research is large. In each small subdivision there are a few specialists who know the background, problems, theories, and facts—who have the knowledge that a system problem may require for its solution. But what small chance is there that one of those few men will meet the system problem? Conversely, those who have the system problems close at hand are system-oriented men, unlikely to be deep in special areas of research. Perhaps each one will know an area or two, but what system project can afford close scrutiny of all of science?

To escape the dilemma we of course appeal to abstracts, handbooks, guides, and journals. The system man hopes to find the needed knowledge, study, and then apply it. Why be an expert in advance if all one needs to know is written? If all *were* written, progress might be made that way, but not enough is written. Compressed sources may convey a few ideas to an expert, but they cannot create an expert. Journal articles are better, but they, too, assume deep background knowledge. No field of science can be mastered in a day—or in a month or year. The logistics of learning thus deny do-it-yourself application.

Then let the system man consult the specialist! Bring face-to-face the one who asks the question and the one who knows the answer. That approach will find solutions, but it also will be slow. One cannot claim solution till a question and its answer both reside in the same mind. So the specialist must learn the system or the system man the science. Either course takes time.

The physical realization of plans comes late in life for modern systems. By testing time, the die is cast, the assembly line is set to run. There is no phase more ravenous for fast decisions, and yet the meaning of a test cannot be found except through deep analysis, which takes much time. Though vast machines transform and smooth the data, their review does not as yet suffice. Before the test can lead to action, there must be an integration that now depends upon a human mind. Once it has the information, the mind may be quick to make decisions, but a bottleneck holds back the information flow. Between the test and the decision lie too many charts, tables, memoranda, reports, and conferences.

In modern systems, as in living organisms, more material and energy go into growth and maintenance than into direct achievement of external goals. Although production and transportation are the more visible facets of logistics, information is the one facet that causes the severest problems. The human hand and eye are much too slow to meet the need for record keeping in major systems. Yet human judgment has not been replaced. The same problem again: we must assimilate the facts before we can achieve control of the logistic situation. But it is

too vast and complex for prompt assimilation through our present methods.

The process of system design must continue throughout system operation. The operational context is ever in flux, the rigid system obsolete before designed. The need is therefore great to keep in hand a detailed picture of the system's present state—to modify, evaluate, and then to modify again.

Once the system is in operation, there may be some chance to make experiments. But when a major system runs, each move can yield a million data, and no two movements are ever quite the same. An operating system is a clinic, not a laboratory. Yet the potential value is so great that it justifies bold efforts to conduct valid research.

THE BASIS FOR A SYSTEM SYSTEM

When many system problems are set forth in close array, one cannot but see the common element, the vast demand for ordered information. Why then should we attack the problems piecemeal? Perhaps there is no reason other than the one now so familiar—the difficulty of mounting a coordinated attack of great proportions. Formidable though that reason be, we cannot let it long forestall concerted effort. The challenge is too great, the need too strong. As major systems grow in size and scope, complexity transcends known methods of control. We must devise new methods to reverse the trend.

The crux of the whole matter lies in the different powers of human beings and present machines that process information. Machines can count and sort in fine detail the prodigious piles of data now required to shape a complex system. Men cannot. Men can sense relevance, see needs, state questions, have hunches, and choose courses. Machines cannot. Neither men nor machines alone can do the job that has been discussed. We must combine the best of each to solve the central system problem. How can we join the widely diverse traits? It will not suffice to subdivide the task into two parts, the fine detail and the overall task, for details by themselves are meaningless and wholes do not exist without their parts.

There is another way to make a team of men and machines. Men work with words, with names and verbs and qualifiers. They tend to talk *about* a complex process. Rarely does any one man execute it. Rarely does he even dictate step by step its execution. When faced with a need for complex action, he delegates the doing to others, who enter when they hear their proper calls. But the others do not by themselves effect it: in turn, they name subprocesses and delegate the doing down

the line. This iterative action cycles on until a twist and turn or two will meet each stated order. A hierarchy of agents may thus be required. In this indirection lies a solution to the problem, for repeated delegation serves to separate the pattern from detail. No matter how complex the whole, at every level all is simple and concise.

The correspondence is direct between this old technique for organizing human efforts and the most modern methods employed in automated information processing. To delegate fulfillment of detail is now termed "to call a subroutine"; to translate the general's directive into detailed orders to the troops is now called "to compile." The advance is not in concept but in speed, precision, and replicability of execution.

The advance gives us, for the first time, a true capability for building giant structures out of microscopic parts. For once a procedure is encapsulated in a subroutine, that procedure will be carried out rapidly and exactly each future time that its name is called. The depth of delegation knows no certain limit. Complex things (or things that at first seem complex) are clarified through subdivision into simple parts. When lower parts are tested and proved right, one shifts attention farther up the scale and finds no more complexity than down below. The strategy is in part to "divide and conquer," in part to weld an empire of vanquished subroutines.

In the process of construction, intuition is as yet the major guide—there is no generic theory. But there are refined techniques. Numerical analysis is full of aids to computation. One need not know the path, only the goal, if he can iterate to a criterion. Recursion sometimes plays a helpful role, achieving deep solutions through repeated use of simple steps.

The language barrier between men and machines has been much noted. It is indeed an imposing obstacle. But there are signs that it is falling. Much can be done with simple words if they are chosen well, precisely understood, and explicitly obeyed. Modern compilers do rapidly and well what most men do poorly—translate from statement of design to procedure that achieves the purpose. Language translation is passing from automated dictionary lookup to automated processing of syntax; machine-translated Russian is not queenly English, but neither is it bad, and 60,000 words a day is rather fast for one translator. In a few fields of research and engineering, specialized languages now bring computers' power into the hands of the experts of those fields, who did not have to desert their specialties to become programmers.

The language problem looms less ominously as more and more procedures are devised and fixed in subroutines. A sentence may then be just a name for a procedure. Or it may state a fact (serve as datum) to be operated upon by other sentences. It may prescribe a mode and thus affect interpretations of other sentences. The organization of elec-

tronic information processing is already so close to the organization of natural language that the discrepancy will not cause great trouble. Some of the differences appear, in fact, to favor computer language over natural language in formulative thinking, problem solving, and control of complex organization.

We have before us, then, as the foregoing notes suggest, a potent tool for shaping systems. This has been proved in scientific calculation, business accounting, and control of industrial processes, although those fields pose relatively simple problems. It has been developed in military command and control and in subsystem simulation, although in a rather rigid way. It has been explored in games, pattern recognition, and problem solving. In the latter fields, application of computers presents both great promise and great difficulty. Those fields involve search for solutions that cannot yet be prescribed through algorithms. Typically the problem spaces are so vast that blind or random search would last for years on fast machines. Heuristic methods, however, may guide the search. In them there is fair ground for hope of real success. But brains are now the sole source of heuristics. Thus the current basic need is that of combining strategic guidance from the human brain with electronics' speed and mammoth memory. By doing so we may create the means with which to organize complex systems.

OUTLINE OF THE SYSTEM SYSTEM

It is difficult to explicate the system that I have in mind. There is no upper bound to force a stop, no lower bound at which one should begin. Perhaps the level at which to start is that of a system aimed to plan, build, and to operate another system. The object system might be a satellite-reconnaissance network or a theater combat operations center. The system system, then, will be the organizing force to make coherent and to harmonize the many facet phases of the network or center.

The system system under examination is one of several or many similar systems. These systems are both enmeshed in suprasystems and conjoined with one another, but let us focus now upon the single entity.

The functions of the system system are representation, communication, problem solving, decision making, and self-adaptation of the object system. The system system has a center and a periphery. At the center is a fast, large-scale computer, somewhat novel in design, specialized to serve many users at the same time. At the periphery are user stations. These range from single computer typewriters to major ganglia with their own computers to mediate their commerce with the rest.

Each man or compact team of men engaged upon the project has typewriter communication with the central computer. Where there is

need, there is provision for faster or more flexible exchange. The facilities make it possible to (1) draw graphs and sketches to be viewed by the computer, (2) receive graphic and pictorial information from the computer, (3) receive fast printouts or reproductions of numerical and textual material from the computer, (4) use the computer to control analog and digital equipment in laboratories and offices, (5) use the computer to arrange communication links and nets (telephone, typewriter, and television) among the men, and (6) use the computer to simulate the object system or its parts, to control war games and tests, and to process data and compute in ordinary ways.

In the computer center there is a large, fast machine of many parts.[1] Instead of a single processing unit there are several. In a given microsecond, most are at work. A few are idle, awaiting calls from consoles or from elements of programs needing action. The many blocks of rapid-access memory respond independently of one another. Each processor interrogates, or stores information into, whichever memory block the program dictates. The number of processors times their speed roughly matches the number of memory blocks times their speed. Therefore not only the processors but also the memory blocks are simultaneously active.

Coordinating the actions of the processors and the memory blocks with one another and with calls from input and output devices are several arbiters. They maintain short waiting lines for the many sequences of program that are in process. No program sequence knows or cares to which processor it will be assigned in the next microsecond, and no processor knows to which block of memory it will next refer. Each unit effects but a single operation; it then turns around to see what its next microtask will be. The assignments are made in accordance with priorities and by a rule that ensures that no unit ever has to wait too long a time.

The fundamental requirement for the time-sharing, assembly-line approach to computation stems from the fact that the computer must attend to many separate needs. Each authorized user must have access to unprespecified information and processes whenever he requires them. Such a demand is radically different from those met by the SAGE system, for example, in which the actions that may be initiated from peripheral consoles are rigidly constrained. One might meet separate needs, of course, by giving each user his own computer, but that would lead away from coordination and toward bankruptcy. It would be almost as unfeasible to assign to each user an individual central processor. The

[1] This is a foreshortened description of a computer concept developed jointly by J. McCarthy, M. Minsky, E. Fredkin, R. L. Silver, W. Fletcher, and myself. We plan in due course to publish a full account of it.

efficient solution involves communal use of central facilities. That solution is best achieved, within the limits of today's technology, in part through parallel processing and in part by rapid switching from one user to another.

The potential value of time sharing is great when a major requirement is that of coordinating the work of many men through written messages. The new IBM Selectric typewriter, for example, can type at an average rate of 1,000 characters per minute. A modest computer (5-μsec memory cycle, one processing unit) can accept a typewritten character and store it in a buffer in about 50 μsec; such a computer can deliver a character to the typewriter in 25. Therefore, if all the computer had to do were to communicate with typewriters, it could handle about 1,600 of them. This figure highlights the key fact that human output and assimilation rates are slow, computers fast. Time sharing is a good device to match them. It lets one spread the cost of a large-scale computer among the many men and yet offer to each almost its full capability.

The main unsharable cost is the cost of random-access memory for each user's personal programs and data. Note, however, that in a large, time-shared system individual programs will be brief and terse. All the basic subroutines, interpreters, and compilers will be public property, shared by many users. (Several users may use a suitably written subroutine simultaneously.) Stored tables of facts and data will also be consulted by several or many users. Communal use will thus make it feasible to have at hand far more complete facilities than any single-user system could afford.

The computer must, of course, serve faster terminal devices than typewriters, and it must execute the many programs which involve input only at the start and those which involve output only at the end. The devices are assigned processors and memory according to their needs. Parallel, multisequence processing may be of as great value to these other operations as to those which involve typewriters. At those moments when a console device demands attention, a processor will provide it in a flash. In the long intervals between demands, however, there is no waste of computer time and cost. This will remove the pressure from programmers to make each individual program efficient in itself; efficiency is inherent in the system.

If special circumstances arise that demand split-microsecond timing, the required number of processors and memory blocks may be withdrawn from the time-sharing pool and devoted solely to the special task. In ordinary work, however, it will seem to any user that he always has the whole machine at his command.

The computer system just described is the hardware core of the system system. Its basic schema permits wide variations in its scale

that make it easy to adapt to system-system needs. In the mission-planning phase and during preliminary design, the project staff will still be small. A full-scale information-processing facility would paralyze, not aid. On the other hand, as many system people know to their regret, starting with a small machine and switching over to a large one in midstream can stop all progress.

The time-sharing computer with multiple user stations, parallel processors, and memory blocks solves the growth problem in a fundamental way. Typewriters, consoles, laboratory terminal units, and other standard and special-purpose equipments may be added as the need for them arises. Each is merely connected to the appropriate input-output bus or buses. As peripheral equipment is added, processors and memory blocks must, of course, be added also.

To integrate new units into the system, it is necessary only to alter a few parameters in the executive routine that monitors and controls the overall operation. Programs that ran when there were two processors and two memory blocks will still run when there are twenty or two hundred. Growth will make it possible to do more things and usually to do things faster, but it will never negate prior gains.

The hardware that has been discussed provides a means for shaping systems, but the guidance and control do not reside within the means. How will men fit into the system system?

Inasmuch as the system system has as its basic purpose the planning, building, and operation of its object system, most of the men of the system system are quite the same men who now plan, build, and operate major systems. In the system system they will work in rather different ways, but their guiding purpose will remain that of the object system.

In every organization, however, there are some who work not to achieve directly the stated system goal but to facilitate the work of others who do. In the system system much such support will be essential. Many men will be required to oversee and to advance the overall facility, to install and repair the means, to write the master programs that will be required, and to advise their object-system-oriented colleagues.

OPERATION OF THE SYSTEM SYSTEM IN VARIOUS PHASES OF THE OBJECT SYSTEM

The computer and its programs will fulfill two related functions of paramount importance, (1) representing the object system and (2) coordinating the efforts of the object system's men. Those functions have long been subserved by blueprints and memoranda, but blueprints

and memoranda cannot meet modern needs. The system system seeks to substitute a way of representing in detail both shape and motion, a way of guiding solution as well as construction, and a way of effecting communication not only rapidly but selectively to those who need the information at the time and in the context of their need.

In the mission-formulation stage, the system system is mainly a war-gaming or management-gaming organization. The chief effort is to determine probable consequences of proposed actions and to find the best course of action. It is not feasible, in such an effort, to incorporate detailed simulations into the games and tests. The interplay among the forces and events within the games and tests must therefore be determined more by decision processes that depend upon assumed relations than by the laws of nature as revealed in the unfolding of analog processes.

The product of the mission-planning phase may be, essentially, a set of high-order routines for a computer program. The high-order routines that constitute the mission set the requirements for detailed action. They present the demands for conditional reaction of the (as yet undefined) system to all the challenges that may be presented by the context within which it will operate. Such a mission formulation is thus like a broad outline in which only major headings are given. It is, however, a dynamic outline, capable of running despite its incompleteness. It is capable of running because all the gaps of substantive knowledge are filled by explicit assumptions. The validity of its performance is conditional upon the correctness of those assumptions, but it nevertheless provides an opportunity to test the overall system concept in advance of actual construction. The approach is also applicable at lower levels of the system plan. It is applied, within the system system, at each new level as that level comes within the focus of attention.

Through such procedures, expanded to include infusion of more complex human judgments, broad plans can be tested far in advance of detailed execution. Although we shall rarely have exact procedures to deduce such things as counterplans or costs and values, that lack should not deter us from being concise and definite about the plans that we do devise. We should not repress from our awareness the uncertainty that made the planning difficult, nor should we transform that uncertainty to vagueness in the plan. We should replace uncertainties with definite assumptions, clearly labeled as assumptions, and revise and adjust the simulation until we have a workable simulated system. Then we must either make good the assumptions or revise them and run the simulation again.

During mission formulation, ideas necessarily arise about means and methods that might be used to carry out the mission plan. Through

gradual transition we now come to a phase in which attention focuses itself upon those means and methods and upon alternatives that we can devise to achieve the broad aims that are now fairly firm.

In this phase, the system system seeks to construct the second stratum of the program that will represent the object system. Free creation is required to synthesize these plans. Yet they must meet requirements set by higher-order routines of the program. These considerations lead to frequent interrogation of the representation. The typewriters, the scope displays, the other features of the time-sharing facility are used in this communication.

Individuals and teams of men thus devise their schemes for fulfilling the formulated mission. Changes of plan at higher levels are at once available to everyone affected by them. Changes still disrupt, but the disruption is minimized. When several alternative system plans have been checked out and proved to run and to fulfill the mission, they are available in clear, dynamic form for everyone concerned to see and judge. No assumptions are hidden. No pet scheme can win acceptance through sheer failure to perceive what is required to make it work.

The product of the competition between alternative designs will be a plan much like a diagram with blocks and arrows, but the plan will have more than two dimensions and one scale. It will unfold its consequences, as stressed before, and make them manifest dynamically in clear displays of process.

As time moves on, attention now descends to still lower strata of the program. The system shape fills out. The project grows in size. The information measure mounts. Though the central focus still remains the representation that we have thus far discussed, other aspects of the picture now come into view—pieces of the object system, schedules, costs, and component tests.

The representation and simulation involved in the early phases were actually prerepresentation and presimulation, for no actual object system then existed. As parts of the object system come into being, however tentatively, their actual, measured characteristics are compared with the characteristics of their prerepresentations. Program changes are made, and the consequences of the changes are examined as the program runs. The ties between the representation and the developing object system are kept close.

Although the developing parts cannot yet be coupled together for subsystem tests, the computer representation functions as a whole. Its operation therefore reveals effects of interactions that without it we could not foresee. Because the men concerned with the conflicting parts communicate with one another through the computer and the communication network that it controls, they can designate precisely to

one another the sources of trouble. They coordinate their actions as they adjust and modify. Never do they let the programmed representation fall behind the actualities that it represents. In many places, of course, it leads and guides. In the others, where actual development has progressed, computer programs are revised with each hardware or personnel advance. Sometimes this process requires several alternative programs for a single function. In such instances, the alternative routines carry explanatory comments ("dynamic footnotes") that present themselves on auxiliary displays as the corresponding processes unfold.

Designers, engineers, and executives are, of course, not the only ones who use computer representation. Stock inventories, procurement schedules, accounting records, vacation plans, and all other facts and figures of the project are in it too. They represent tremendous tasks, but tasks of the same magnitude must even now be handled. The difference lies in availability of the facts upon a moment's notice and in the circumstance that in the computer the facts can be related to one another with the aid of megacycle clerks.

In every phase of the development, there is a need to bring to bear upon the growing object system the best existing knowledge. It has been noted how hard it is to meet that need. Suppose, however, that research results are published in two ways. (1) A journal note describes the problem and the nature of the research performed to solve it. It presents a broad summary of the results and discusses implications. It presents graphs, but never detailed tables. (2) A computer program provides a precise, dynamic model of the process studied. It is both a theory and a set of facts, for it contains research results set into relation with one another and with their context. The problem itself is indexed or "descriptorized." Information about the program format, calling sequences, and the like, is contained in a machine-readable preamble.

Application of some research results would then involve only: (1) *Retrieval*, in which the system man locates the research with the aid of the index and the information-processing machines. (The system man, of course, obtains the research report and programs and reads the former.) (2) *Matching*, in which parameters of the research problem are compared with parameters of the system representation to determine whether or not the research answers a relevant question. (3) *Format translation*, in which, if the research model is relevant, the program is converted to a format compatible with the rest of the system representation. (Through standardization, format translation is made a machine process.) (4) *Incorporation*, in which the research model is incorporated into the system representation and the system program (or a part of it that includes the research model) is run.

The significance of the research results for the object system reveals

itself through dynamic displays of system operation. Parameters of the research model may be varied in a search for optimal system performance. Alternative models may be compared. New approaches may be suggested by this study.

The advance achieved by the system system is to transfer the complex model in full detail from research to application, bypassing human brains except in highest-order matters. That solves the problems posed by informational complexity and by the slowness of man's assimilation of detailed information.

From development to test we have, again, a gradual transition. The focus shifts to exercises of the object system as a whole or in large segments. These exercises, however, are paralleled by exercises of the representation, and every object test result is cross-compared against a like test of the program. When any test result demands a change, that change is studied first in simulation. If the problem can be solved in simulation, all the requirements set by the solution are then introduced into the programmed plan.

Schedules of tests and changes are themselves designed by the same method as that used in the design of the object system. A simulation runs to check the feasibility of each plan before it is made official. All those whom it affects learn of it through the computer and its communication net. Each schedule change is there related to all programmed plans, and any plan ceases to operate successfully in the system context as soon as it is outmoded by an alteration in system schedule or design.

Presimulation of each system test prepares the way for rapid analysis of actual test data. The problem at each stage is essentially that of finding the cause of every discrepancy between actual performance and performance of the model. In finding the cause, as in reducing data, the computer facility provides the basic tools.

The movement from prototype to production brings about a major reorientation of the system system. Its aims are still the solution of problems and coordination, but the problems and the actions change in nature. Those parts of the system system concerned with schedules and inventories grow. Time and funds become the dominant concerns.

Many production operations are facilitated by the existence of precise plans in program form. Referring to programmed models, the computer now controls milling machines and lathes. It operates the lines on which electronic and mechanical subsystems are assembled. It counts each part and product and compares the progress with the schedule.

Wherever human operations are involved in production, the computer serves both as teacher and as supervisor. It can prescribe suc-

cessive manipulations through headphones and through displays at every workplace. It can detect flaws by making tests as soon as each unit is assembled. Whatever can be routinized it can perform or check, and most steps in production are routine.

Operation of the object system, we may safely assume, is expensive. It may also be slow or infrequent. Indeed, if the system is of a military nature, it may never be unleashed to prove itself, yet it must remain forever ready. These factors interfere with a major function of operation—to continue system tests and thereby provide a steady feedback of successes and failures. This testing function must be served because the context within which the object system operates will not stand constant, nor will it be the same context for which the early plans were made. If the object system fails to adapt to its changing context, it rapidly becomes obsolete, and early obsolescence is rarely advantageous in large systems.

The system-system simulation must therefore be retained as part of the operating scheme. Indeed the operators may run the simulation more than they run the object system. With the simulation they explore new techniques and tactics. The best they then may employ in actual runs, which serve as critical experiments.

One may introduce simulation directly into actual operations. An operator need never know which mode is current. When he is working with simulated signals, his performance can be trained, for the computer always knows which signals come from targets and which are merely noise. The computer can provide the operator with the knowledge of results—the reinforcement—required to advance his skill. Teams may be trained in this way, of course, as well as individuals. The range of tasks to which this scheme applies is great. It includes not only target detection (in which it has been most widely exploited) but also decision making, pattern recognition, problem solving, and the other complex functions. In operational scheduling, logistics, and maintenance, also, the system system plays its coordinating role. It endows the object system with self-awareness and with the capacity to reflect and plan. The system system thus constitutes an overlay. It is a part of what it planned and built.

The system system is too grandiose a scheme, I fully realize, to propose or to predict without explicit expression of belief that such things must see the light of day in this real world. Such things must come to pass, in my opinion, or shortly we shall bury ourselves in uncontrolled complexity.

APPENDIXES

Appendix I[1]

SELECTED LIST OF HANDBOOKS, TEXTBOOKS, AND GENERAL REFERENCES PERTINENT TO HUMAN FACTORS ENGINEERING

1. Abruzzi, A.: "Work Measurement: New Principles and Procedures," Columbia University Press, New York, 1952.

2. Abt, L. E.: Human Engineering, *Ann. N.Y. Acad. Sci.,* vol. 51, pp. 1123–1276, 1951.

3. Allen, Patricia S., and Ezra V. Saul: An Annotated Bibliography of Bibliographies Pertinent to the Design and Use of Machines by Human Operators, *Human Factors,* vol. 1, pp. 26–44, September, 1958.

4. Armsby, D. H., and K. G. Cook: Design Standards for Man-Machine Tasks in Signal Corps Systems, Contract DA 36 039 SC 78328, Project 3 99 00 110, Applied Psychology Corporation, Arlington, Va., December, 1959.

5. Armstrong, H. G.: "Principles and Practice of Aviation Medicine," The Williams & Wilkins Company, Baltimore, 1952.

6. Baker, Charles A., and Walter F. Grether: Visual Presentation of Information, *USAF WADC Tech. Rept.* 54-160, August, 1954.

7. Bamford, Harold E., et al.: The Operation of Manned Spacecraft, *USAF WADC Tech. Rept.* 58-225, June, 1958.

8. Banghart, F. W., and E. G. Pattishall: Human Factors at Extreme Altitudes: Synopsis and Bibliography, *USAF ARDC Tech. Rept.* 60-7, Contract AF 18(600) 1792, March, 1960.

9. ——— and ———: Human Factors at Extreme Altitudes: Synopsis and Bibliography II, *USAF ARDC Tech. Rept.* 60-7, Contract AF 18(600) 1792, May, 1960.

10. Barnes, R. M.: "Motion and Time Study," 3d ed., John Wiley & Sons, Inc., New York, 1949.

11. ———: "Work Measurement Manual," 4th ed., Wm. C. Brown Co., Dubuque, Iowa, 1951.

12. ———: "Work Methods Training Manual," 3d ed., Wm. C. Brown Co., Dubuque, Iowa, 1951.

13. Bartley, S. H., and Eloise Chute: "Fatigue and Impairment in Man," McGraw-Hill Book Company, Inc., New York, 1947.

[1] Contributed by P. G. Ronco, D. B. Devoe, and E. V. Saul.

14. Beranek, L. L.: "Acoustics," McGraw-Hill Book Company, Inc., New York, 1954.

15. Berger, C., H. C. Reese, and C. A. Feder: A Psychological Operations Bibliography, Special Operations Research Office, American University, Washington, May, 1960.

16. Bevan, William, and Rollin N. Patton: Selected Bibliography: Fatigue, Stress, Bodily Change and Behavior, *USAF WADC Tech. Rept.* 57–125, April, 1957.

17. A Bibliography of Aerospace Medical Division Reports in the Fields of Engineering Psychology and Training Psychology, Behavioral Science Laboratory, U.S. Air Force Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, 1960.

18. Bibliography on Key Set Design, International Business Machines Corporation Advanced Systems Development Laboratory, San Jose, Calif., Mar. 15, 1961.

19. Bills, A. G.: "The Psychology of Efficiency," Harper & Brothers, New York, 1943.

20. Black, J. W., and W. E. Moore: "Speech," McGraw-Hill Book Company, Inc., New York, 1955.

21. Bloomfield, L.: "Language," Henry Holt and Company, Inc., New York, 1948.

22. Bolt, R. H., et al.: Handbook of Acoustic Noise Control, vol. 1, Physical Acoustics, *USAF WADC Tech. Rept.* 52-204, December, 1952.

23. Bond, H. J., and R. B. Sleight: Human Factors in the Design of Desert Equipment, Final Report, DA Project 595-20-001, ORD Project TB1-1000, Aberdeen Proving Ground, Md., December, 1954.

24. Bradley, James V.: Glove Characteristics Influencing Control Manipulability, *USAF WADC Tech. Rept.* 57-389, August, 1957.

25. Bray, C. W.: "Psychology and Military Proficiency," Princeton University Press, Princeton, N.J., 1948.

26. Burns, N. M., R. B. Ziegler, and E. C. Gifford: The Effects of Long Term Confinement on Perception, Personality and Performance, U.S. Navy Air Crew Equipment Laboratory, Naval Air Materiel Command, Philadelphia, May, 1960.

27. ——— et al.: Environmental Requirements of Sealed Cabins for Space and Orbital Flights: A Bibliography of Psychophysiological Studies Relevant to Space and Orbital Flight, *Naval Air Materiel Command Air Crew Equipment Lab. Rept.* 441, October, 1960.

28. Chapanis, A.: The Design and Conduct of Human Engineering Studies, *San Diego State Coll. Foundation Tech. Rept.* 14, July 18, 1956.

29. ———: Human Engineering, in C. D. Flagle et al. (eds.), "Operations Research and Systems Engineering," pp. 534–582, Johns Hopkins Press, Baltimore, 1960.

30. ———: "Research Techniques in Human Engineering," Johns Hopkins Press, Baltimore, 1959.

31. ———: On Some Relations between Human Engineering, Operations Research and Systems Engineering, *Johns Hopkins Univ. Psychol. Lab. Rept.* 8, May, 1960.

32. ———, W. R. Garner, and C. T. Morgan: "Applied Experimental Psychology: Human Factors in Engineering Design," John Wiley & Sons, Inc., New York, 1949.

33. Checklist of Human Engineering Design Principles, Missile Division, North American Aviation, Inc., Los Angeles, January, 1959.

34. Christensen, J. M.: Trends in Human Factors, *Human Factors,* vol. 1, no. 1, pp. 2–7, 1958.

35. Churchill, Edmund, Alma Kuby, and D. S. Daniels: Nomograph of the Hand and Its Related Dimensions, *USAF WADC Tech. Rept.* 57-198, April, 1957.

36. Clendening, L.: "The Human Body," 4th ed., Alfred A. Knopf, Inc., New York, 1949.

37. Conover, Donald W.: Study of the Human Element in Future Antiballistic Missile Systems, *Convair Human Factors Eng. Group Tech. Rept.* ZG-017-31, December, 1960.

38. Cybernetics and Man, English translations from *Priroda, Moscow,* Order 61-11013 from Office of Technical Services, Department of Commerce, Washington.

39. Debeln, W. V., and L. Laurell: Anthropometric Data and Cabin Space, Meddelanden fran flygoch Navalmedicinska Namnden, Stockholm, 1960. (In Swedish.)

40. Dempster, W. T.: The Anthropometry of Body Action, *Ann. N.Y. Acad. Sci.,* vol. 63, pp. 559–585, 1955.

41. ———: Space Requirements of the Seated Operator. Geometrical, Kinematic, and Mechanical Aspects of the Body with Special Reference to the Limbs, *USAF WADC Tech. Rept.* 55-159, July, 1955.

42. ———, W. C. Gabel, and W. J. L. Felts: The Anthropometry of the Manual Work Space for the Seated Subject, *Am. J. Phys. Anthropol.,* vol. 17, no. 4, pp. 289–317, 1959.

43. Design for Maintainability: Final Management Report, *USN Training Device Center Tech. Rept.* 330-1, September, 1959.

44. Design for Maintainability, Technical Supplement I: An Experimental Investigation of Equipment Packaging for Ease of Maintenance, *USN Training Device Center Tech. Rept.* 330-2, June, 1959.

45. Design for Maintainability, Technical Supplement II: Survey of Information, (*A*) Description of the Training Device Technician, and (*B*) Engineering Estimates of Circuits and System Maintainability, *USN Training Device Center Tech. Rept.* 330-3, July, 1959.

46. Development of an Index of Electronic Maintainability: A Research Report, American Institute for Research, Pittsburgh, June, 1959.

47. Dreyfus, H.: "Designing for People," Simon and Schuster, Inc., New York, 1955.

48. Dunlap, J. W.: Men and Machines, *J. Appl. Psychol.,* vol. 31, no. 6, pp. 565–579, 1947.

49. Eckenrode, R. T.: The Response of Man to His Environment, Dunlap and Associates, Inc., Stamford, Conn., 1958.

50. Ely, J. H., R. M. Thomson, and J. Orlansky: Design of Controls, *USAF WADC Tech. Rept.* 56-172, November, 1956.

51. ———, ———, and ———: Layout of Workplaces, *USAF WADC Tech. Rept.* 56-171, September, 1956.

52. Emanuel, Irvin, John W. Chaffee, and John Wing: A Study of Human Weight Lifting Capabilities for Loading Ammunition into the F-86H Aircraft, *USAF WADC Tech. Rept.* 56-367, August, 1956.

53. ——— et al.: A Height-Weight Sizing System for Flight Clothing, *USAF WADC Tech. Rept.* 56-365, April, 1959.

54. Ernst, Arthur A.: Feasibility Study for a Man-Machine Systems Research Facility, *USAF WADC Tech. Rept.* 59-51, March, 1959.

55. Evans, P. M.: "An Introduction to Color," John Wiley & Sons, Inc., New York, 1948.

56. Evans, R. N., and L. J. Smith: A Study of Performance Measures of Trouble Shooting Ability on Electronic Equipment, Contract N60RI 17042, Project NR 153 124, College of Education, University of Illinois, Urbana, Ill., October, 1953.

57. Evrard, E., P. Bergeret, and P. M. Van Wulfften Palthe (eds.): "Medical Aspects of Flight Safety (The Unexplained Aircraft Accident)," Pergamon Press, Inc., New York, 1959.

58. Factors Influencing the Maintenance of Electronic Equipment, U.S. Naval Training Device Center, Port Washington, N.Y., January, 1958.

59. Fatigue and Stress Symposium, January 24-26, 1952, *Operations Research Office Tech. Mem.* ORO-T-185, September, 1952.

60. Fisher, F. R.: Protection and Function of the Hands in Cold Climates, National Academy of Sciences, National Research Council, Washington, 1957.

61. Fitts, P. M.: Human Factors in Air Navigation and Traffic Control Systems, Civil Aeronautics Administration, National Research Council, Washington, 1949.

62. ———: Notes and Selected Readings on Human Engineering Concepts and Theory, College of Engineering, Summer Session, University of Michigan, Ann Arbor, Mich., August, 1959.

63. ———: Psychological Research on Equipment Design in the AAF, *Am. Psychologist*, vol. 2, no. 3, pp. 93–98, 1947.

64. Flaherty, Bernard E. (ed.): "Psychophysiological Aspects of Space Flight," Columbia University Press, New York, 1961.

65. Fletcher, H.: "Speech and Hearing in Communication," 2d ed., D. Van Nostrand Company, Inc., Princeton, N.J., 1953.

66. Flight Safety Foundation, Inc.: The Psychologist as Member of the Design Team, *Human Factors Bull.*, no. 60-3H, 1960.

67. Flight Simulator Utilization Handbook, *Human Factors Operations Research Lab. Rept.* 42, August, 1953.

68. Floyd, W. F., and D. F. Roberts: Anatomical Physiological and Anthropometric Principles in the Design of Office Chairs and Tables, *Brit. Standards Inst. Rept.* BS 3044, 1958.

69. ——— and A. T. Welford (eds.): "Symposium on Fatigue," H. K. Lewis, London, 1953.

70. ——— and ———: "Symposium on Human Factors in Equipment Design," H. K. Lewis, London, 1954.

71. Folley, J. D., and J. W. Altman: Guide to Design of Electronic Equipment for Maintainability, *USAF WADC Tech. Rept.* 56-218, April, 1956.

72. ——— and ———: Human Factors in Maintenance—Part IV. Factors Influencing the Maintenance of Electronic Equipment, *USN Training Device Center Tech. Rept.* NAVTRADEVCEN 20-OS-23-4, Sept. 22, 1958.

73. Ford, Henry, II: The Challenge of Human Engineering, *Advanced Management,* vol. 2, no. 2, pp. 48–51, 1946.

74. Fryer, D. H., and E. R. Henry: "Handbook of Applied Psychology," Rinehart & Company, Inc., New York, 1950.

75. Fundamentals of Design of Piloted Aircraft Flight Control Systems, vol. 3, The Human Pilot, *Bur. Aeronaut. Rept.* AE-61-4, August, 1954.

76. Gagne, R. M., and E. A. Fleishman: "Psychology and Human Performance," Henry Holt and Company, Inc., New York, 1959.

77. Geldard, Frank: Engineering Psychology: A British Psychological Society Symposium, *Occupational Psychol.,* vol. 31, no. 4, pp. 209–217, October, 1957.

78. ———: "The Human Senses," John Wiley & Sons, Inc., New York, 1953.

79. Glassner, H. F., and G. A. Peters: Bio-electronic Analysis of Performance, *Douglas Aircraft Co. Eng. Paper* 897, March, 1960.

80. ——— and ———: Bio-electronic Analysis of Performance, *Douglas Aircraft Co. Eng. Paper* 898, May, 1960.

81. Goldman, Alexander: Sensory Interaction and Response Capacity, *USN Training Device Center Tech. Rept.* NAVTRADEVCEN 789-11-1, June, 1959.

82. Grant, G., and J. M. McKendry: Design for Maintainability, *USN Training Device Center Tech. Rept.* NAVTRADEVCEN 330-1, April, 1960.

83. Gregg, Lee W.: Changes in Muscular Tension during Psychomotor Performance, *USA Quartermaster Research Eng. Center Environmental Protection Research Div. Tech. Rept.* EP-54, May, 1957.

84. Hanrahan, J. S., and D. Bushnell: "Space Biology: The Human Factors in Space Flight," Basic Books, Inc., New York, 1960.

85. Hansen, Robert, Douglas Y. Cornog, and H. L. Yoh: Annotated Bibliography of Applied Physical Anthropology in Human Engineering, *USAF WADC Tech. Rept.* 56-30, May, 1958.

86. Hasbrook, A. Howard: Human Factors—The Basis for Crash Safety Design, *Flight Safety Foundation, Inc., Aviation Crash Injury Research Rept.* 63-0-108, September, 1959.

87. "Heating, Ventilating and Air Conditioning Guide," vol. 24, American Society of Heating and Ventilation Engineers, New York, 1956.

88. Helson, H.: Design of Equipment and Optimal Human Operation, *Am. J. Psychol.,* vol. 62, pp. 473–497, 1949.

89. Hertzberg, H. T. E., and G. S. Daniels: Anthropometry of Flying Personnel, *USAF WADC Tech. Rept.* 52-321, 1954.

90. Hick, W. E., and J. A. V. Bates: The Human Operator of Control Mechanisms, *Ministry Supply Mon.* 17-204, London, May, 1950.

91. Hooten, E. A., et al.: A Survey in Seating, Heywood-Wakefield, Gardner, Mass., 1945.

92. Hopkins, C. O., D. K. Bauerschmidt, and M. J. Anderson: Display and Control Requirements for Manned Space Flight, *USAF WADD Tech. Rept.* 60-197, April, 1960.

93. Human Engineering Design Standards for Missile Systems Equipment (and Supplements), *USAF Ballistic Missile Div. Exhibit* 57-8A, Nov. 1, 1958.

94. Human Factors in Accidents, American Institute for Research, Pittsburgh, 1950. (Restricted.)

95. Hunt, M. M.: Doctor for Push Buttons, *Nation's Business*, vol. 40, no. 4, pp. 48, 52–56, April, 1952.

96. "Illuminating Engineering Society Lighting Handbook: The Standard Lighting Guide," 2d ed., Illuminating Engineering Society, New York, 1952.

97. Javitz, A. E. (ed.): Human Engineering in Equipment Design, *Elec. Mfg.*, vol. 58, no. 4, pp. 107–126, October, 1956.

98. ———: Introduction to Human Engineering in Product Design, *Elec. Mfg.*, vol. 49, no. 3, pp. 90–95, March, 1952.

99. Jones, C. E., J. L. Kobrick, and H. F. Gaydos: Anthropometric and Biochemical Characteristics of the Hand, *USA Quartermaster Research Eng. Center Environmental Protection Research Div. Tech. Rept.* EP-100, September, 1958.

100. Jones, E. M., R. H. Gaylord, and J. D. Folley, Jr.: Guide to Human Engineering of Miniaturized Equipment. Final Report, *Am. Inst. Research Rept.* AIR-261-59-FR-202, June, 1959.

101. Judd, D. B.: "Color in Business, Science, and Industry," John Wiley & Sons, Inc., New York, 1952.

102. King, Barry G.: Functional Cockpit Design, paper presented at Civil Aeronautics Administration, Institute of Aeronautical Sciences Annual Meeting, New York, 1952.

103. ———: Measurements of Man for Making Machinery, *Am. J. Phys. Anthropol.*, vol. 6, no. 3, p. 341, 1948.

104. ———, D. J. Morrow, and E. P. Vollmer: Cockpit Studies. The Boundaries of the Maximum Area for the Operation of Manual Controls, *USN Med. Research Inst. Rept.* 3, 1947.

105. Kobrick, John L.: Spatial Dimensions of the 95th Percentile Arctic Soldier, *USA Quartermaster Research Eng. Center Environmental Protection Research Div. Tech. Rept.* EP-39, September, 1956.

106. Kryter, Karl D.: Human Engineering Principles for the Design of Speech Communication Systems, *USAF CCDD Tech. Rept.* 60-27, August, 1960.

107. Lerner, H. D., H. J. Bond, and R. B. Sleight: The Relation of Toxic Gases to Equipment Design, Project 595-20-001, Human Engineering Laboratory, Aberdeen Proving Ground, Md., 1954.

108. "Lighting Handbook," Illuminating Engineering Society, New York, 1954.

109. Losee, J. E., and J. A. Buongiorno: Maintainability and Support-

ability Evaluation Technique, pt. I, *USAF WADD Tech. Note* 60-82, March, 1960.

110. Luckiesh, M.: Seeing, Lighting Research Laboratory, General Electric Company, Cleveland, Ohio, no date.

111. Lukasik, S. J., and A. W. Nolle (eds.): Handbook of Acoustic Noise Control, vol. 1, Physical Acoustics, suppl. 1, *USAF WADC Tech. Rept.* 52-204, April, 1955.

112. McCormick, E. J.: "Human Engineering," McGraw-Hill Book Company, Inc., New York, 1957.

113. McFarland, R. A.: The Application of Human Body Size Data to Vehicular Design, *SAE Spec. Publ.* 142, 1955.

114. ———: "Human Factors in Air Transport Design," McGraw-Hill Book Company, Inc., New York, 1946.

115. ———: "Human Factors in Air Transportation: Occupational Health and Safety," McGraw-Hill Book Company, Inc., New York, 1953.

116. ———, R. C. Moore, and A. B. Warren: Human Variables in Motor Vehicle Accidents: A Review of Literature, Harvard School of Public Health, Boston, 1955.

117. ——— et al.: Human Body Size and Capabilities in the Design and Operation of Vehicular Equipment, Harvard School of Public Health, Boston, 1953.

118. ——— et al.: Human Factors in the Design of Highway Transport Equipment: A Summary Report of Vehicle Evaluation, Harvard School of Public Health, Boston, 1953.

119. ——— et al.: Proceedings President's Conference on Industrial Safety, U.S. Department of Labor, March, 1949.

120. McGinnis, John: Handbook of Criteria Used by Troops in Evaluating Quartermaster Clothing and Personal Equipment, *USA Quartermaster Research Eng. Center Environmental Research Protection Div. Tech. Rept.* EP-74, December, 1957.

121. McGrath, James J., Albert Harabedian, and Donald N. Buckner: Human Factors Problems in Anti-submarine Warfare. Review and Critique of the Literature on Vigilance Performance, Human Factors Research, Inc., Los Angeles, December, 1959.

122. McGrath, Joseph E., and Peter G. Nordlie: Research Methodology of Requirement-setting Studies, Human Sciences Research, Inc., Arlington, Va., November, 1959.

123. McKendry, J. M., et al.: Maintainability Handbook for Electronic Equipment Design, suppl. 4 to NAVTRADEVCEN 330-1, Design for Maintainability, *USN Training Device Center Rept.* NAVTRADEVCEN 330-1-4, April, 1960.

124. Mackworth, N. H.: Research on the Measurement of Human Performance, *Med. Research Council Spec. Rept.* 268, 1950.

125. Magoun, F. A.: "Problems in Human Engineering," The Macmillan Company, New York, 1932.

126. Maintainability of Electronic Equipment, vol. 1, *Proc. First Electronics Industries Assoc. Conf.*, 1958.

127. Maintainability of Electronic Equipment, vol. 2, *Proc. Second Electronics Industries Assoc. Conf.*, 1959.

128. Maintainability of Electronic Equipment, vol. 3, *Proc. Third Electronics Industries Assoc. Conf.*, 1960.

129. Marsh-Thesis, Inc.: Voice Intelligibility Handbook, *USN Special Devices Center Rept.* NAVEXOS P-1192, August, 1953.

130. Mathematical Models of Human Behavior: Proceedings of a Symposium, Dunlap and Associates, Inc., Stamford, Conn., 1955.

131. Mead, L. C.: A Program in Human Engineering, *Ann N.Y. Acad. Sci.*, vol. 51, pp. 1123–1278, 1951.

132. Melton, A. W., and G. E. Briggs: Engineering Psychology, *Ann. Rev. Psychol.*, vol. 11, pp. 71–98, 1960.

133. Meltzer, H.: Review of Reviews in Industrial Psychology, 1950–1959, *Personnel Psychol.*, vol. 13, no. 1, pp. 31–58, Spring, 1960.

134. Miller, G. A.: "Language and Communication," McGraw-Hill Book Company, Inc., New York, 1951.

135. Miller, R. B.: Handbook on Training and Training Equipment Design, *USAF WADC Tech. Rept.* 53-136, June, 1953.

136. ———: Human Engineering Design Schedule for Training Equipment, *USAF WADC Tech. Rept.* 53-138, June, 1953.

137. ——— and J. D. Folley, Jr.: Recommendations on Designing Electronics Equipment for Job Maintenance, *Human Resources Research Center Research Bull.* 51-33, December, 1951.

138. Miner, R. W. (ed.): Dynamic Anthropometry, *Ann. N.Y. Acad. Sci.*, vol. 63, pp. 433–636, 1955.

139. Morehouse, L. E., and A. T. Miller: "Physiology of Exercise," The C. V. Mosby Company, St. Louis, 1948.

140. Morgan, C. T., and E. Stellar: "Physiological Psychology," 2d ed., McGraw-Hill Book Company, Inc., New York, 1950.

141. Morse, P. M., and G. E. Kimball: "Methods of Operations Research," The Technology Press, Cambridge, Mass., and John Wiley & Sons, Inc., New York, 1956.

142. Muckler, F. A., et al.: Psychological Variables in the Design of Flight Simulators for Training, *USAF WADC Tech. Rept.* 56-369, January, 1959.

143. Nall, M. L.: Submarine Habitability: A Literature Survey (1951–1959), A., Atmospheric Control, B., Performance under Stress, Contract Nonr 2718(00), Man-Machine Information Center, Documentation Incorporated, Washington, May, 1959.

144. ———: Working Bibliography: Noise Levels in Aircraft, Contract Nonr 2718(00), Man-Machine Information Center, Documentation Incorporated, Washington, April, 1959.

145. Naval Electronics Laboratory Reliability Design Handbook, PB 121839, Office of Technical Services, Department of Commerce, 1954.

146. Nordlie, Peter G.: Methodology for Analysis of Man's Role in an Advanced Space Flight System, Human Sciences Research, Inc., Arlington, Va., November, 1959.

147. Nutt, A. B.: Principles of Canopy Design for Military Aircraft, *USAF WADC Tech. Rept.* 53-92, April, 1953.

148. O'Connor, J.: "Born That Way," The Williams & Wilkins Company, Baltimore, 1928.

149. ———: "Psychometrics," Harvard University Press, Cambridge, Mass., 1934.

150. Optical Society of America, Committee on Colorimetry: "The Science of Color," Thomas Y. Crowell Company, New York, 1953.

151. Peters, G. A.: A Guide to Display of Information, *Control Eng.*, vol. 7, no. 5, p. 123, May, 1960.

152. ——— and H. F. Glassner: Trends in Applied Bio-mechanics, *Douglas Aircraft Co. Eng. Paper* 972, February, 1960.

153. Pieron, H.: "The Sensations: Their Functions, Processes, and Mechanisms," Yale University Press, New Haven, 1953.

154. Pletcher, K. E.: Human Factors in Aerospace Pathology, *Aerospace Med.*, vol. 32, no. 1, pp. 6–11, January, 1961.

155. Procedure Plan Human Factors Methodology, Contract DA 36 039 SC 80078, Ramo-Wooldridge Corporation, Los Angeles, September, 1959.

156. Proceedings of Symposium on Decision Theory and Applications to Electronic Equipment Design, *USAF RADC Tech. Rept.* 60-70A, April, 1960.

157. Quastler, H. (ed.): "Information Theory in Psychology," Free Press, Glencoe, Ill., 1955.

158. Raben, Margaret W.: A Survey of Operations and Systems Research Literature, Contract Nonr 494(13), Institute for Applied Experimental Psychology, Tufts University, Medford, Mass., January, 1960.

159. Randall, F. E., A. Damon, and R. S. Benton: Human Body Size in Military Aircraft and Personal Equipment, *AAF Tech. Rept.* 5501, 1946.

160. Recommendations on Designing Electronic Equipment for the Job of Maintenance, *Human Resources Research Center Research Bull.* 51-33, December, 1951.

161. Ritchie, Malcolm L., and Charles A. Baker: Psychological Aspects of Cockpit Design—A Symposium Report, *USAF WADC Tech. Rept.* 57-117, April, 1957.

162. Rohles, Frederick, H., Jr. (ed.): A Preliminary Survey of Human Factors Personnel, Aeromedical Field Laboratory, Holloman Air Force Base, N.Mex., January, 1960.

163. Rosenblith, W. A., and K. N. Stevens: Handbook of Acoustic Noise Control, vol. 2, Noise and Man, *USAF WADC Tech. Rept.* 52-204, June, 1953.

164. Rowland, G. E., and C. M. Kulp: A Method of Making Dimensional Measurements of Complex Motions, *Rowland & Co. Rept.* 60-1-2, March, 1960.

165. Sells, S. B., and C. A. Berry (eds.): "Human Factors in Jet and Space Travel," The Ronald Press Company, New York, 1961.

166. Shapero, A., et al.: Human Engineering Testing and Malfunction

Data Collection in Weapon System Test Programs, *USAF WADD Tech. Rept.* 60-36, February, 1960.

167. Silver, C. A.: Documentation Index and Bibliography for Study on Human Engineering of Control Systems, *Minneapolis-Honeywell Regulator Co. Aero Rept.* 1508 TR1, November, 1959.

168. Slechta, R. F., et al: Comparative Evaluation of Aircraft Seating Accommodation, *USAF WADC Tech. Rept.* 57-136, April, 1957.

169. Smith, Bertram J. (ed): Human Control Dynamics Air and Space Craft, Flight Safety Foundation, Inc., New York, November, 1958.

170. Spector, W. S. (ed.): Handbook of Biological Data, *USAF WADC Tech. Rept.* 56-273, October, 1956.

171. Standards to Be Employed in Research on Visual Displays, Armed Forces–National Research Council Vision Committee, Ann Arbor, Mich., March, 1950.

172. Stevens, S. S. (ed.): "Handbook of Experimental Psychology," John Wiley & Sons, Inc., New York, 1951.

173. Suggestions for Improving the Maintainability of Military Electronic Equipment for the Designer, Installer, User and Maintainer, Electronics Industries Association, New York, 1957.

174. "Survey Report on Human Factors in Undersea Warfare," National Research Council, Washington, 1949.

175. "Survey Report on Human Factors in Undersea Warfare, Supplement," National Research Council, Washington, 1954.

176. Survey of Research Projects in the Field of Aviation Safety, Guggenheim Aviation Safety Center, Cornell University, New York, 1951.

177. Survey of Research Projects in the Field of Aviation Safety: First Annual Supplement, Guggenheim Aviation Safety Center, Cornell University, New York, 1952.

178. Taylor, C. L.: Environmental Biotechnology Syllabus for Engineering 130, Department of Engineering, University of California, Los Angeles, February, 1951.

179. ——— and L. M. K. Boelter: Biotechnology: A New Fundamental in the Training of Engineers, *Science*, vol. 105, pp. 217–219, 1947.

180. Taylor, F. V.: Four Basic Ideas in Engineering Psychology, *Am. Psychologist*, vol. 15, pp. 643–649, 1960.

181. ———: Psychology and the Design of Machines, *Am. Psychologist*, vol. 12, pp. 249–258, 1957.

182. Thomson, Robert M., et al.: Arrangement of Groups of Men and Machines, *ONR Rept.* ACR-33, December, 1958.

183. Tiffin, J.: "Industrial Psychology," 3d ed., Prentice-Hall, Inc., Englewood Cliffs, N.J., 1952.

184. Uhlaner, J. E.: Systems Research—Opportunity and Challenge for the Measurement Research Psychologist, *USA Personnel Research Branch Tech. Research Note* 108, July, 1960.

185. Van Cott, Harold P., and James W. Altman: Procedures for Including Human Engineering Factors in the Development of Weapons Systems, *USAF WADC Tech. Rept.* 56-488, October, 1956.

186. Vandenberg, J. D.: Facts about Human Engineering, *Advanced Management,* vol. 18, pp. 25–27, November, 1953.

187. ——— and C. Thomas Goldsmith: Human Factors Engineering, *Machine Design,* vol. 30, nos. 4, 5, 8–10, 12, 13, Apr. 17, 1958, pp. 108–113, May 1, 1958, pp. 104–109, May 15, 1958, pp. 114–118, June 12, 1958, pp. 123–126, June 26, 1958, pp. 123–126, July 10, 1958, pp. 109–113.

188. Verdier, Paul A.: "Basic Human Factors for Engineers," Exposition Press, New York, 1960.

189. Wallis, D.: Human Factors in Design and Use of Naval Equipment, Royal Naval Research Committee, Medical Research Council, London, February, 1960.

190. Wever, E. G., and M. Lawrence: "Physiological Acoustics," Princeton University Press, Princeton, N.J., 1954.

191. Weybrew, B. B.: Bibliography of Sensory Deprivation, Isolation and Confinement, *USN Med. Research Lab. Mem. Rept.* 60-1, January, 1960.

192. Whittenburg, John A.: Methodology for Evaluation of a Man-Machine Surveillance System, Human Sciences Research, Inc., Arlington, Va., December, 1959.

193. Williams, R. H. (ed.): Human Factors in Military Operations, *Dept. Army Tech. Mem.* ORO-T-259, January, 1954.

194. Wilson, W. S.: Man-Machine Space Complex, *Aircraft and Missiles,* vol. 3, pp. 22–26, July, 1960.

195. Wood, C. C.: Human Factors Engineering, *Ergonomics,* vol. 1, pp. 294–300, 1958.

196. Wood, F. W.: A Design Guide to Selection and Application of Knobs, *Machine Design,* vol. 32, no. 11, pp. 118–120, May 26, 1960.

197. Wright, G. O.: A General Procedure for Systems Study, *USAF WADD Tech. Note* 60-18, January, 1960.

198. Wulfeck, J. W., A. Weisz, and M. W. Raben: Vision in Military Aviation, *USAF WADC Tech. Rept.* 58-399, November, 1958.

199. Yaglou, C. P.: Thermal Standards in Industry, *Am. J. Public Health,* vol. 40, no. 5, pp. 131–143, May, 1950.

200. Yarnold, K. F., and R. C. Channel: Systems Research with Special Reference to Human Engineering, *ONR Rept.* 641-2-13, January, 1952.

201. Zergin, Robert S., Edmund Churchill, and M. Alexander: A Head Circumference Sizing System for Helmet Design, *USAF WADC Tech. Rept.* 60-631, December, 1960.

202. Ziegler, R., and J. Lazo: Medical and Human Engineering Aspects to Flight in Non-Conventional Aircraft, USN Air Crew Equipment Laboratory, Naval Air Materiel Command, Philadelphia, October, 1959.

Appendix II[1]

ABRIDGED TOPICAL OUTLINE OF THE LITERATURE IN HUMAN FACTORS ENGINEERING

1.0.0 HUMAN ENGINEERING: METHODS, FACILITIES, EQUIPMENT AND GENERAL REFERENCES

- 1.1.0 Bibliographies and general references
- 1.2.0 Methods and design procedures
  - 1.2.1 Statistical methods
  - 1.2.2 Methods of task and personnel description
  - 1.2.3 Psychophysical methods
  - 1.2.4 Physiological methods
  - 1.2.5 Special techniques
- 1.3.0 General equipment and apparatus
- 1.4.0 Facilities in human engineering

2.0.0 SYSTEMS OF MEN AND MACHINES

- 2.1.0 Bibliographies and general references
- 2.2.0 Design of systems and operations
  - 2.2.1 Communication and information theory
  - 2.2.2 Game or decision theory and linear programming
  - 2.2.3 Computers and simulation
  - 2.2.4 Queueing theory and work measurement techniques
- 2.3.0 Research and evaluation of systems
  - 2.3.1 Assignment of functions to men or machines
  - 2.3.2 Groups as system components
  - 2.3.3 Communications systems
  - 2.3.4 Transportation systems
  - 2.3.5 Production maintenance and supply systems
  - 2.3.6 Air traffic control system

[1] Contributed by P. G. Ronco, D. B. Devoe, and E. V. Saul.

3.0.0 VISUAL INPUTS AND PROCESSES
- 3.1.0 Bibliographies and general references
- 3.2.0 Natural ambient lighting
  - 3.2.1 Daytime light
  - 3.2.2 Twilight and night
  - 3.2.3 Special conditions affecting visibility
  - 3.2.4 Glare
- 3.3.0 Artificial ambient lighting
  - 3.3.1 Considerations of illumination
  - 3.3.2 Lighting systems, outdoor
  - 3.3.3 Lighting systems, indoor
  - 3.3.4 Illumination, unusual characteristics
- 3.4.0 Lighting of instruments
  - 3.4.1 Direct lighting and floodlighting
  - 3.4.2 Indirect lighting
  - 3.4.3 Color and intensity of illumination
  - 3.4.4 Comparisons of methods and types
- 3.5.0 Radarscopes and other cathode-ray displays
  - 3.5.1 Signal detectability
  - 3.5.2 Range and bearing scales and aids
  - 3.5.3 Size, shape, lighting, etc., of screen
- 3.6.0 Television and motion picture displays
- 3.7.0 Pictorial and symbolic displays
  - 3.7.1 Outside-in and inside-out displays
  - 3.7.2 Combining pictorial and symbolic display elements
  - 3.7.3 Comparisons among types of displays
- 3.8.0 Indicators and scales
  - 3.8.1 Counters
  - 3.8.2 Pointers
  - 3.8.3 Scales: shape, size, and direction of increase
  - 3.8.4 Scales: divisions and markings
  - 3.8.5 Design of scales for qualitative readings
  - 3.8.6 Evaluation and comparison of indicators and scales
- 3.9.0 Legibility of letters, numerals and other symbolic forms
  - 3.9.1 Design of characters
  - 3.9.2 Color and contrast between symbol and background
  - 3.9.3 Viewing conditions
- 3.10.0 Printed materials
  - 3.10.1 Graphs and tables
  - 3.10.2 Maps and charts
  - 3.10.3 Decals, check lists, instruction charts, etc.
  - 3.10.4 Evaluation and comparison of types of printed materials
  - 3.10.5 Photography and photo interpretation
- 3.11.0 Camouflage or concealment
- 3.12.0 Visual coding and visual search
  - 3.12.1 Object characteristics

3.12.2 Light coding
3.13.0 Optical aids
3.13.1 Devices for visual enhancement
3.13.2 Protective devices
3.14.0 Other factors affecting visual performance
3.15.0 Basic visual data related to the design and use of equipment
3.15.1 Individual differences and anomalies
3.15.2 Threshold visibility
3.15.3 Adaptation, pre-adaptation, and pre-exposure
3.15.4 Perception of color
3.15.5 Brightness discrimination
3.15.6 Acuity
3.15.7 Special effects dependent upon fixation or exposure time
3.15.8 Eye movements
3.15.9 Perception of depth, distance, and size
3.15.10 Perception of form, contour, and pattern
3.15.11 Perception of number, angle, and direction
3.15.12 Perception of movement
3.16.0 Equipment and methods for basic and applied problems in vision
3.16.1 Tests of color vision
3.16.2 Other tests of visual performance
3.16.3 Equipment and methods for basic visual research problems
3.16.4 Simulators and tests for specific applied problems

4.0.0 AUDITORY INPUT AND PROCESSES, INCLUDING SPEECH PRODUCTION AND INTELLIGIBILITY

4.1.0 Bibliographies and general references
4.2.0 Ambient noise
4.2.1 Measurement of noise level and composition
4.2.2 Noise reduction and control
4.2.3 General industrial and equipment noise
4.2.4 Aircraft and weapons noise
4.2.5 Submarine and ship noise
4.2.6 Effects of ambient noise and blast on performance
4.2.7 Noise-induced hearing loss
4.3.0 Effects of auditory equipment components
4.3.1 Input devices
4.3.2 Transmission devices
4.3.3 Output devices
4.4.0 Evaluations of specific systems in speech communication
4.4.1 Telephone and intercom systems
4.4.2 Radio systems
4.5.0 Evaluations of specific systems in nonverbal auditory display
4.5.1 Intermittent warning and signaling devices
4.5.2 Telegraphic systems
4.5.3 Sonar and other underwater sound systems
4.5.4 Flybar

4.6.0 Characteristics of auditory signals in relation to coding
4.7.0 Special auditory skills
4.8.0 Basic data in the production and perception of speech
- 4.8.1 Basic characteristics of speech
- 4.8.2 Speech audiometry and articulation testing
- 4.8.3 Speech masking and the signal-to-noise ratio
- 4.8.4 Speech distortion
- 4.8.5 Individual differences and anomalies
- 4.8.6 Language design
- For Training in voice communication—see 14.1.0
- 4.8.7 Synthetic speech

4.9.0 Basic data in audition
- 4.9.1 Basic attributes: pitch
- 4.9.2 Basic attributes: loudness
- 4.9.3 Basic attributes: timbre, duration, etc.
- 4.9.4 Thresholds and related phenomena
- 4.9.5 Aftereffects of stimulation
- 4.9.6 Stimulus mixture
- 4.9.7 Sound localization
- 4.9.8 Auditory patterns and meaning
- 4.9.9 Psychological scaling
- 4.9.10 Norms, individual differences, and anomalies
- For Nonverbal auditory training—see 14.1.0
- 4.9.11 Physiological mechanisms
- 4.9.12 Equipment and methods in research in audition and speech

5.0.0 OTHER SENSORY INPUTS AND PROCESSES

5.1.0 Touch
- 5.1.1 Basic processes and data
- 5.1.2 Tactile coding
- 5.1.3 Vibratory and electrical stimuli used as signals and displays
- 5.1.4 Equipment and methods used in HE research on touch

5.2.0 Temperature sensitivity
- 5.2.1 Basic processes and data
- 5.2.2 Equipment and methods used in HE research on temperature sensitivity

5.3.0 Pain
- 5.3.1 Basic processes and data
- 5.3.2 Equipment and methods used in HE research in pain

5.4.0 Smell and taste
- 5.4.1 Basic processes and data
- 5.4.2 Olfactory and gustatory signals
- 5.4.3 Equipment and methods used in HE research on smell and taste

5.5.0 Kinesthesis
- 5.5.1 Basic processes and data
- 5.5.2 Coding and signalling through kinesthesis
- 5.5.3 Equipment and methods used in HE research on kinesthesis

5.6.0 Vestibular functions
5.6.1 Basic processes and data
5.6.2 Equipment and methods used in HE research on vestibular functions
5.7.0 Time perception

6.0.0 INPUT CHANNELS: CHOICE AND INTERACTION

6.1.0 Bibliographies and general references
6.2.0 Comparison of input channels
6.2.1 Comparisons of visual and auditory channels
6.2.2 Comparisons of channels other than visual and auditory
6.3.0 Intersensory effects
6.3.1 Facilitation and inhibition of reception
6.3.2 Factors determining orientation in space
For Sensory deprivation—see 12.8.0

7.0.0 BODY MEASUREMENTS, BASIC PHYSIOLOGICAL CAPACITIES, BASIC AND COMPLEX MOTOR PERFORMANCE

7.1.0 Bibliographies and general references
7.2.0 Anthropometric measurements
7.2.1 Body size, stationary
7.2.2 Body size in motion or in unusual positions
7.3.0 Body mechanics
7.3.1 Extent of limb movement
7.3.2 Flexibility of movement
7.3.3 Muscular strength and endurance
7.4.0 Equipment and methods used in HE research on anthropometry
7.5.0 Basic physiological capacities
7.5.1 Equipment and methods in research on basic physiological capacities
7.6.0 Basic motor performance
7.6.1 Positioning movements
7.6.2 Repetitive and rhythmic movements
7.6.3 Manual dexterity
7.6.4 Reaction time
7.6.5 Handedness
7.6.6 Involuntary reflexes
7.6.7 Equipment and methods in HE research on basic motor performance
7.7.0 Complex motor performance
7.7.1 Watchkeeping performance
7.7.2 Tracking performance
7.7.3 Serial performance
7.7.4 Equipment and methods in HE research on complex motor performance

8.0.0 DESIGN OF CONTROLS AND INTEGRATION WITH DISPLAYS

8.1.0 Bibliographies and general references
8.2.0 Standardization of controls
8.3.0 Types of controls
8.3.1 Rotary movement controls
8.3.2 Linear movement controls
8.3.3 Other types of controls
8.3.4 Multifunction controls: combined controls
8.3.5 Comparisons among types of controls: choice of type of control
8.4.0 Display-control coding
8.4.1 Multiple dimensions
8.4.2 Labelling
8.5.0 Positioning and plane of operation of controls relative to operator
8.6.0 Special considerations relating to prolonged adjustment
8.7.0 Control dynamics
8.7.1 Display-control movement ratios
8.7.2 Control loading
For Human reaction times—see 7.6.4
8.7.3 Compatibility
For Tracking—see 7.7.2
8.7.4 Aided controls
8.7.5 Quickened displays

9.0.0 LAYOUT OF PANELS AND CONSOLES

9.1.0 Bibliographies and general references
9.2.0 Standardization and integration of panels and consoles
9.3.0 Location of panel relative to operator and task
9.4.0 Grouping of components on panels and consoles
9.4.1 Ease of discrimination
9.4.2 Spatial dynamics
9.5.0 Specific orientation of parts

10.0.0 DESIGN OF WORK SPACE, EQUIPMENT, AND FURNITURE

10.1.0 Bibliographies and general references
10.2.0 Workplace design
10.2.1 Visibility
10.2.2 Ease and speed of movements
10.2.3 Stowage
10.3.0 Furniture specifications
10.3.1 Seating and body support
10.3.2 Seating arrangements
10.3.3 Work surfaces
10.4.0 Passageways, entrances, exits
10.5.0 Tools
10.6.0 Design for complex motor coordination tasks and special body positions

10.7.0 Design for maintenance
10.8.0 Design for portability in the design of equipment
10.9.0 Design for safety
10.9.1 Motor vehicle safety
10.9.2 Air safety
10.10.0 Specific work places and equipments not elaborated below
10.10.1 Cockpits, space cabins, capsules
10.10.2 Air- and spacecraft
10.10.3 Sea and landcraft
10.10.4 Fire control (weapons) systems and equipment
10.10.5 Industrial equipment

11.0.0 CLOTHING AND PERSONAL EQUIPMENT

11.1.0 Bibliographies and general references
11.2.0 Clothing ensembles
11.2.1 Thermal protection
11.2.2 Pressure suits
11.2.3 Other types of protective clothing
11.2.4 Fabrics for clothing
11.3.0 Clothing components
11.3.1 Belting
11.3.2 Body gear
11.3.3 Headgear
11.3.4 Handgear
11.3.5 Footgear
11.4.0 Clothing size
11.5.0 Personal equipment
11.5.1 Auditory devices
For Visual protective devices—see 3.13.2
11.5.2 Sleeping bags
11.5.3 Packs and carriers
11.5.4 Parachutes, life jackets, life rafts, and other survival equipment
11.5.5 Prosthetics
11.6.0 Effects of combinations of clothing and personal equipment
11.7.0 Shelters
11.8.0 Equipment and methods used in HE research on clothing and personal equipment

12.0.0 SPECIAL ENVIRONMENTAL FACTORS AFFECTING PERFORMANCE

12.1.0 Bibliographies and general references
12.2.0 Thermal environment
12.2.1 Temperature, humidity, and air velocity
12.2.2 Thermal radiation
12.3.0 Toxic environments
12.4.0 Motion
12.4.1 Speed and acceleration

12.4.2 Vibration
12.4.3 Motion sickness
12.5.0 Altitude and depth
12.5.1 Atmospheric pressure
12.5.2 Oxygen requirements
12.6.0 Nuclear and cosmic radiation
12.7.0 Space travel
12.8.0 Sensory deprivation
For Factors determining orientation in space—see 6.3.2
For Unusual characteristics of artificial ambient lighting affecting visual performance—see 3.3.4
For Effects of ambient noise and blast on performance—see 4.2.6
12.9.0 Special equipment and methods utilized in the study of the effects of special environments on performance

13.0.0 INDIVIDUAL FACTORS, WORK CONDITIONS, AND TASK CHARACTERISTICS THAT AFFECT BEHAVIORAL EFFICIENCY

13.1.0 Bibliographies and general references
13.2.0 Factors pertaining to the individual
13.2.1 Motivation and emotion
13.2.2 Intelligence and aptitudes
13.2.3 Thought processes
13.2.4 Attention, alertness, vigilance
13.3.0 Interactions between individual factors and work factors
13.3.1 Effects of individual understanding of task or set toward task
13.3.2 Acceptability of equipment, clothing, food, and/or task
13.3.3 Fatigue and behavior decrement
13.3.4 Stress
13.4.0 Work, rest, and efficiency
13.4.1 Conditions of work
13.4.2 Methods of work
13.4.3 Levels of complexity
13.4.4 Unusual characteristics of work
13.5.0 Performance as affected by certain physiological processes and selected dietary and pharmacological substances
13.5.1 Sleep
13.5.2 Diet, food, and nutrition
13.5.3 Drugs
13.5.4 Aging

14.0.0 TRAINING AIDS AND DEVICES AND THEIR USE

14.1.0 Bibliographies and general references

15.0.0 OTHER AREAS OF PSYCHOLOGICAL RESEARCH PERTINENT TO HUMAN ENGINEERING

15.1.0 Personnel psychology relevant to HE
15.2.0 Social psychology relevant to HE

INDEX

Date Due

Due	Returned	Due	Returned
SEP 12 1988	AUG 22 1988		